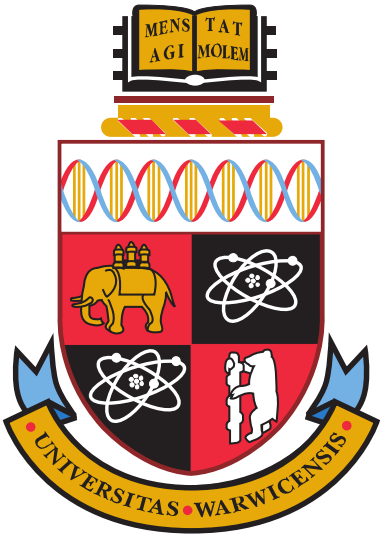


Accurate Light and Colour reproduction in High Dynamic Range Video Compression

by

Ratnajit Mukherjee

Thesis

Submitted to the University of Warwick

for the degree of

Doctor of Philosophy

Warwick Manufacturing Group

January 2017

THE UNIVERSITY OF WARWICK

Contents

List of Tables

List of Figures

Acknowledgments

Firstly, I would like to thank my supervisors Professor Alan Chalmers and Dr. Kurt Debattista for providing me the opportunity to do a PhD. Their continuous support, advice, encouragement and enthusiasm are sincerely appreciated. While Kurt was a great mentor who patiently guided me throughout my PhD, I am very thankful to Alan for not only providing me the opportunity to pursue a PhD but also the opportunity to participate as a working group member of European Union Cost Action IC1005, where I had the opportunity to meet and work with some of the leading academics of the field.

My sincerest gratitude to Dr. Rafal Mantiuk, Dr. Peter Vangorp, Dr. Maximino Bessa and Dr. Miguel Melo for their continous support, guidance and cooperation which led to a very good publication and one of the most important chapters of my thesis.

My sincerest gratitude to the members of Visualisation group whose constant support, advice and help during the tough times (acadmic or otherwise) is greatly appreciated. I would like to extend a special thanks to Elmedin for guiding me during the early phases on my PhD. Your advice during those days helped me throughout my PhD. I am grateful to Tom and Carlo for their constant guidance and (very) critical outlook which helped me to improve my perception towards research.

My sincerest gratitude and thanks to Debmalya, Sayan, Partha and Argha for supporting me in every single way, when I needed the most. My special thanks to Irene for sharing all the good and bad times with me and exhibiting that persistence even in the darkest of times allows us to achieve our goals. My sincerest gratitude to my family for their constant encouragement and inspiration without which it would have been difficult to do a PhD, five thousand miles away from home.

Finally, thanks to HPC Cluster Wales for their computational resources, University of Stuttgart and Technicolor for providing some of the contents used in this PhD, EPSRC and Jaguar Landrover for funding this research.

Declarations

The work in this thesis is original and no portion of this work has been submitted in support of an application for another degree or qualification at this university or at another university or institution of learning.

Abstract

HIGH Dynamic Range (HDR) imaging has the potential to replace traditional Low Dynamic Range (LDR) imaging due to HDR's capability to accurately capture and reproduce the entire spectrum of visible lighting conditions with full colourimetric precision in any scene. However, this ability comes at the cost of significantly increased storage and transmission requirements compared to traditional LDR imaging. These costs together with additional challenges in capturing and delivering HDR video, for example ghosting artefacts, peak luminance of current HDR displays etc., are currently limiting the faster adoption of HDR imagery and the eventual replacement of traditional LDR imaging and video techniques.

This thesis focuses on how to deliver high-fidelity HDR video with minimal storage/transmission requirements. To answer such a multi-faceted question, the thesis first provides an overview of HDR imaging and video pipeline followed by a detailed discussion on existing HDR video compression algorithms and quality assessment (QA) techniques for HDR image and video. This background information provides an in-depth review of the overall progress made to date and also highlights the current outstanding issues.

The thesis subsequently assesses end-user preference of HDR video content over LDR video content using a rating- and a ranking-based psychophysical experiment. Results from this assessment suggest that there exists a statistically significant difference between the HDR representation of a scene and its LDR counterparts where given the option, the former is preferred by end-users as HDR provides a more realistic viewing experience.

Having established the preference for HDR video, a comprehensive objective and subjective study is undertaken of a number of published/patented HDR video compression algorithms by means of several objective QA metrics and psychophysical studies. This resulted in an in-depth understanding of the advantages and shortcomings of existing solutions. Results obtained demonstrate that *non-backward* compatible compression algorithms are able to deliver high-fidelity HDR video at significantly lower storage/transmission costs compared to *backward* compatible algorithms. Also, perceptual QA metrics exhibit a high to very high correlation with subjective video quality assessment.

Based on this in-depth understanding of the design requirements and philosophy of HDR video compression algorithms, this thesis proposes and evaluates a novel HDR video compression algorithm. This new algorithm is shown to outperform existing state-of-the-art algorithms both in terms of image reconstruction quality and transmission requirements.

Chapter 1

Introduction

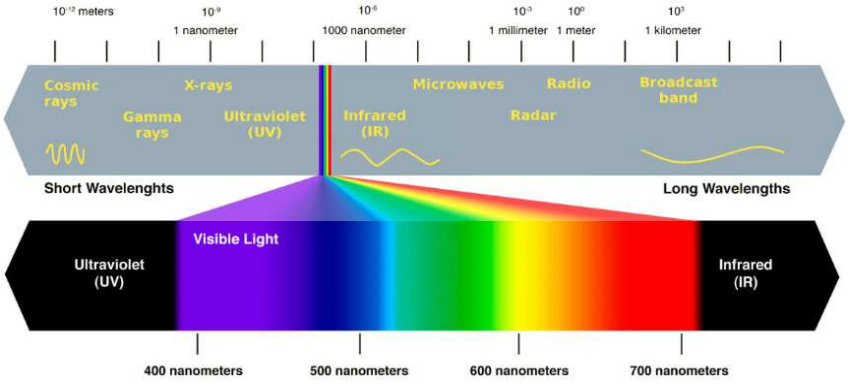
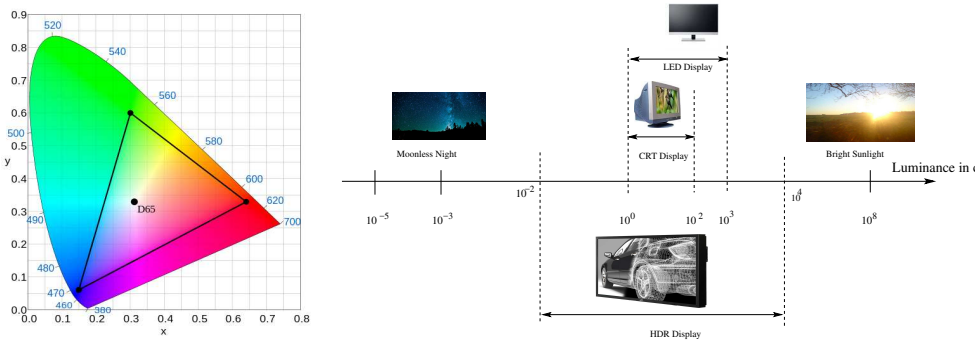


Figure 1.1: Schematic diagram of electromagnetic energy spectrum - with projected visible spectrum. Picture courtesy *Digiolighting*.

LIGHT is a form of electromagnetic energy which travels in space and interacts with materials where it can be absorbed, reflected, refracted or transmitted. This electromagnetic energy is a continuous spectrum where the wavelength can range from 10^{-14} metres (high-energy gamma rays) to 10^{6} metres (low energy radio waves). The light visible to the human eye is a fraction of this continuous spectrum. The wavelength of the visible light ranges from $\approx 380-780$ nm as shown in Figure 1.1. In spite of the limited energy range visible to the human eye, the human visual system (HVS) is remarkable in being able to adapt and cope with a wide range of visible lighting conditions in the real world [BADC11]. The HVS is capable of adapting to environmental lighting conditions that vary approximately by 13 orders of magnitude ranging from starlit night ($\approx 10^{-5}$ cd/m^2) to bright daylight ($\approx 10^{8}$ cd/m^2). In comparison, the majority of existing digital imaging systems, also referred to as Low Dynamic range (LDR) or Standard Dynamic Range (SDR) imaging systems, can capture and display only a fraction (up to three orders of magnitude at most) of the light visible to the HVS.



(a) REC. 709 colour gamut (area covered by triangle) of traditional display devices with D65 white-point.

(b) Luminance chart displaying the dynamic range covered by traditional and HDR display devices

Figure 1.2: Colour gamut and luminance chart displaying the capabilities of the HVS compared to display capabilities of traditional and HDR displays.

1.1 Limitations of LDR imaging

The limitation of LDR imaging can largely be attributed to the limitations of the capture, storage (file formats) and display technologies. Most LDR cameras are limited in terms of the colour gamut (see Figure 1.2a) and dynamic range (contrast), that they can capture. Furthermore, a majority of LDR imaging and video file formats such as Joint Photography Experts Group (JPEG) and Motion Pictures Experts Group (MPEG) offer up to 8 bits/pixel/channel i.e. 24 bits/pixel (bpp) encoding of the input light stimulus. Finally, display technologies such as cathode ray tube (CRT) displays and liquid crystal displays (LCDs) provide a peak luminance of $\approx 80-350$ cd/m^2. However, such an assumption of device referred data is no longer true, since newer generations of image acquisition devices (cameras) and displays are able to capture and depict a wider colour gamut and dynamic range than their CRT and LCD predecessors. For instance, existing high-end digital cameras provide an extended dynamic range for static images with proprietary file formats such as *.raw, .cr2, .nef* and *.pef*, encoding up to 14 bits/pixel/channel or 42 bpp, while state-of-the-art light emission displays (LEDs) offer a peak luminance levels of $\approx 500-1000$ cd/m^2. Therefore, traditional imaging, confined to 8-bits/pixel/channel i.e 24 bpp integer representation, cannot offer the precision required by upcoming imaging technologies, which attempt to match the capabilities of the HVS.

1.2 A general introduction to High Dynamic Range imaging

High Dynamic Range (HDR) imaging brings a complete paradigm shift in digital imaging by being able to overcome the limitations of traditional imaging. HDR imaging can capture and encode the entire dynamic range, as seen by the HVS at a specific (ambient light)

adaptation level, with full colourimetric precision. Also, HDR file formats offer up to 32 bits/channel i.e. 96 bpp encoding, which matches and potentially exceeds the capabilities of the HVS [RHD*10]. Pixel values in HDR are specified by a triplet of floating point matrices, where each matrix represents a colour channel. The floating point data representation enables HDR to represent real world (physical) luminance values and is sometimes referred to as *scene-referred* data.

1.2.1 Applications of HDR

The pioneering works in this field were the creation of HDR file formats, extended colour spaces and native HDR displays [1] which proved that the visualisation of real world colour and luminance ranges were possible. One of the earliest adopters of the precision of HDR imaging were game developers for game engines along with graphics card developers. Most state-of-the-art game engines perform rendering using HDR imaging and deliver more appealing imagery by subsequently tone-mapping [2] the rendered HDR images for commercial (LDR) displays. In addition, special effect production houses now deliver more believable imagery using the precision provided by HDR imaging. State-of-the-art cinema cameras already capture significantly higher dynamic range and colour details than displayable by most commercial displays. Cinematographers and photographers working with high contrast scenes, capture and manipulate scene details (previously unavailable) by using HDR imaging techniques. Apart from commercial applications, HDR imaging can significantly improve scientific applications such as medical imaging (the DICOM [MDG08] standard) or computer vision [Sze10] due to its capability to capture and process more information than existing LDR imaging techniques.

1.2.2 Limitations of HDR imaging

Although HDR imaging has significant advantages over traditional imaging, it comes with its own set of major challenges which need to be addressed before it can be widely adopted for commercial and scientific purposes. Foremost among these is the data storage issue. Due to its floating point precision, HDR content is considerably larger in size, compared to LDR data. A single full-HD HDR image with the resolution of 1920×1080 pixels and full-floating point precision can take up to 24 MB of storage space. Therefore, it is quite evident that the storage issue is more acute while capturing HDR video content at 24/30 frames/second (fps), where a large amount of floating point data needs to be stored in real-time. Also, video codecs are a vital component in the compression process and state-of-the-art codecs do not support native HDR encoding to date. Furthermore, even if HDR content is generated, stored and encoded, currently it can only be decoded and played

[1] see Chapter 2 for details.

[2] an overview of tone-mapping is introduced later in Chapter 2

back using tailor-made HDR image/video players on native HDR displays (discussed later in Chapters 2) as most commercial displays are unable to provide native HDR support.

1.2.3 HDR video compression

Given that a single uncompressed full-HD HDR video frame can take up to 24 MB of storage space, an uncompressed HDR video of 60 seconds duration, can take up to $24 \times 30 \times 60$ MB i.e. ≈ 42 GB of storage space, assuming a capture rate of 30 fps. Therefore, it is quite evident that such storage/transmission requirements are impractical for real-world applications.

Unlike the notion of traditional video compression, *HDR video compression* refers to the *pre-processing* of native HDR video data such that the HDR content is converted to a format suitable for video encoders. Correspondingly, the encoded HDR video stream(s) undergo *post-processing* for reconstruction of HDR video frames. Although a considerable body of research has been conducted on proposing and benchmarking HDR video compression algorithms, the lack of a single well-defined open standard (currently a number of standards exist) to compress HDR video content has resulted in an eclectic collection of solutions.

This thesis focuses on HDR video compression and brings together a comprehensive literature review on several aspects of HDR video compression such as colour space transformations, perception based transfer functions and HDR image/video quality evaluation techniques used for compression related purposes. In addition to these concepts, the thesis verifies the user preference of HDR video over LDR video by means of subjective evaluation techniques, provides a comprehensive literature review of existing HDR video compression algorithms, an in-depth understanding of the design decisions taken to create the existing algorithms and an evaluation of these by means of objective and subjective techniques. Finally, based on the knowledge gained from these evaluations including the advantages and shortcomings of existing algorithms, the thesis answers the following primary research question.

1.3 Research question

The primary research question that this thesis aims to answer is: *What are the design decisions required to deliver high-fidelity High Dynamic Range (HDR) video at minimal storage and transmission cost ?*

To answer this research question, this research will first answer a few additional fundamental questions. They are as follows:

1. *Is HDR video really preferred over LDR video purely from a viewers' perspective with/without a contextual narrative?* This is fundamentally important since the enormous infrastructure, research and development effort required to produce and deliver

HDR content for commercial purposes might be of little importance if HDR fails to deliver the enhanced viewing or cinematic experience (only visual)[3] that has been advertised and assumed to date.

2. *What are the existing algorithms by which HDR video content can be compressed and delivered at feasible storage and transmission costs?* Only an in-depth understanding of the design decisions taken to create such compression algorithms and a comprehensive evaluation of the same can lead to the thorough understanding of the advantages and shortcomings of existing HDR compression and delivery mechanisms.

3. Based on the knowledge gained by answering the previous two questions: *Is a novel HDR video compression algorithm needed to overcome the limitations of the existing state-of-the-art and what other design parameters such as state-of-the-art image/video processing techniques need to added/changed/included to deliver superior video reconstruction quality at lower storage/transmission costs than existing state-of-the-art?*

1.4 Research methodology

A visual description of the research methodology is given in Figure 1.3. As discussed in Section 1.3, the shift from traditional LDR imaging to HDR imaging requires an enormous infrastructure, research and development effort, and monetary investment. Since the scope of scientific applications is minimal compared to the mainstream entertainment, it is fundamentally important to verify whether HDR video is indeed preferred over LDR video purely from the viewers' perspective. Such a verification work can only be conducted by means of a subjective experiment and should be considered as an important preliminary step before further research in HDR video based applications. HDR video can be mapped to its LDR counterpart via a number of different techniques other than tone-mapping operators. Such mapping techniques ensure that the viewers are presented with a choice of techniques which were originally designed for alternative purposes and in turn can lead to a different viewing experience. To conduct this verification work, several HDR videos can be shortlisted and mapped to their corresponding LDR versions using different mapping techniques where each technique is a representative of a class of HDR to LDR mapping techniques. Subsequently, multiple subjective evaluations (for additional verification) need to be conducted where the participants can be tasked to rate/rank the reference HDR videos and their corresponding LDR versions in order of their viewing preference. If multiple evaluations are conducted for additional verification, it is preferable to have mutually exclusive group

[3]Enhanced cinematic experience also consists of auditory as well as 3-D viewing experience which is out of scope of this thesis.

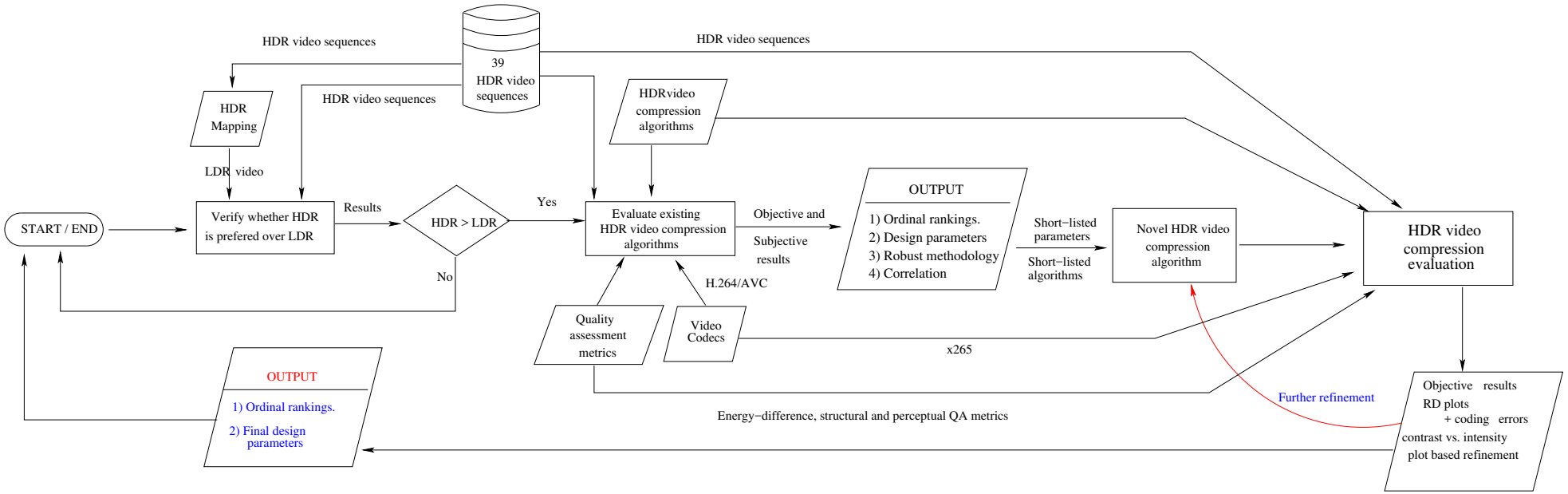


Figure 1.3: Schematic diagram of the research methodology

of participants since it eliminates participant bias. Furthermore, the sequence and order in which the videos are presented to the participants should be random thereby ensuring unbiased quality assessment. If the results obtained from the subjective evaluations suggest that there exists a statistically significant difference between the HDR representation of a scene and its corresponding LDR representations, with the viewers preferring the former, then the first sub-question discussed in Section 1.3 is satisfactorily answered and the research can proceed further to answer the next question. However, certain constraints need to be noted. Amongst the plethora of HDR to LDR mapping techniques, it is possible that the mapping techniques chosen for such a work might not always be the optimal candidates. Also, the HDR video sequences for such a study are typically not part of any contextual narrative upon which the user preference might change. The details of this work is discussed in Chapter 5.

If the clear preference of HDR videos over LDR videos is established the next step is to identify and comprehensively evaluate the existing HDR video compression algorithms. This can be considered as the first step towards answering the primary research question mentioned in Section 1.3. A thorough and fair evaluation requires a common framework where existing HDR video compression algorithms are re-implemented from the original research papers. In addition, the pre-processed HDR video sequences need to be compressed using the same codec at multiple quality levels and reconstructed HDR video sequences should be evaluated against the same set of objective metrics for a fair comparison. The results obtained from the objective evaluation can then be used to benchmark the performance of the candidate algorithms using rate distortion (quality versus output bitrate) plots. Additional subjective evaluations are required to reverify the objective results and draw a correlation between the objective and subjective results. Such a detailed study of existing HDR video compression algorithms provides an in-depth understanding of the advantages and disadvantages of each HDR video compression approach as well as several underlying concepts such as colour space transformation, usage of perceptual transfer functions, secondary luminance streams and their effects on the transmission rate and reproduction quality of HDR videos. The combination of the objective and subjective results can be used to identify the performance based ordinal rankings of the candidate algorithms. In addition, the correlation between the objective and subjective evaluation can also be used to identify and short-list the most accurate objective assessment techniques. Finally, such a thorough objective and subjective evaluation provides a detailed and robust methodology which can be followed for future evaluations. The details of this work is described in Chapter 6 and the objective evaluation framework in Appendix A.

Finally, the in-depth understanding of the design decisions (including advantages and shortcomings) behind each existing algorithm can be used to short-list the design decisions and parameters required for efficient HDR video compression. Based on the short-listed parameters, a novel HDR video compression algorithm can be proposed which ad-

dresses the short-comings of existing solutions to deliver superior HDR video quality at lower storage and transmission costs. The proposed algorithm can then be evaluated against existing state-of-the art, based on literature as well as the best algorithms identified from the previous evaluation, using a similar methodology as described in Chapter 6 using the framework described in Appendix A. The evaluation results can be used for further refinement of the proposed algorithm until it is able to deliver better HDR reconstruction at lower transmission rate compared to existing algorithms. The details of this work is described in Chapter 7.

1.5 Main contributions

Parts of this thesis have been published as papers in two journals [MDBR*16b, MDBR*16a]. These publications form the core of this thesis.

The contributions of this research are:

1. A verification of the superiority and preference of HDR video over existing LDR video, purely from a viewer's perspective. This work is a video specific extension of the work done by Akyuz et al. [AFR*07]. It confirms the fundamental assumption that HDR images are preferred over their LDR counterparts also holds true for HDR videos in spite of the introduction of many efficient, sophisticated and perceptually accurate state-of-the-art HDR to LDR mapping techniques. This work establishes that the acceptance of HDR video is not limited to scientific applications and given the correct viewing conditions naïve users are more likely to prefer HDR video content than existing LDR video content. This conclusion is verified by the rating- and ranking-based subjective evaluations (with 28 and 27 users, respectively) results where rating scores and ranking orders establish the preference of HDR videos over their LDR counterparts. The results show that overall there exists a statistically significant difference between the HDR representation of a video and its corresponding LDR representations. Therefore, the investment of time and effort required to establish an end-to-end HDR video pipeline could lead to a more widespread adoption of HDR in the mainstream media.

2. A comprehensive evaluation of six state-of-the-art HDR video compression algorithms using 39 HDR video sequences, seven full-reference (where both the reference and target (to be evaluated) images are available) objective quality assessment metrics and two ranking-based subjective evaluations. It provides an in-depth understanding of the overall HDR video compression schema and the factors responsible for efficient compression of HDR videos. This work establishes the ordinal ranking of each algorithm based on their performance thereby identifying the best performing algorithm amongst the chosen six. Results suggest that *non-backward* compatible

algorithms deliver superior video quality at lower transmission costs compared to *backward* compatible algorithms. Furthermore, a correlation between objective and subjective evaluation results demonstrate than perception based QA metrics have a high to very high correlation with subjective evaluation results and can thus be used to reliably benchmark HDR video compression algorithms. Additionally, the work also provides a robust methodology for evaluating compression algorithms in the future.

3. A novel HDR video compression algorithm which delivers better HDR video reconstruction quality compared to state-of-the-art solutions. This is achieved by using IPT colour opponent space for effective decorrelation and manipulation of intensity and chroma components, a new analytical transfer function to perceptually encode the brightness information and a novel error minimisation scheme to non-linearly encode the colour information. Additionally, the modular design of the proposed algorithm facilitates the usage of any existing transfer function to non-linearly encode the intensity information.

1.6 Thesis organisation

This thesis is organised as follows:

- Chapter 2 provides the necessary background information on HDR imaging, HDR pipeline, file formats, colourimetric description and colour spaces frequently used for HDR image processing and tone-mapping.

- Chapter 3 focuses on HDR video compression and provides the necessary background on HDR video compression approaches and transfer functions. It describes the state-of-the-art HDR video compression algorithms in detail; some of which are used throughout this thesis and finally it provides an overview of the codecs used for video compression.

- Chapter 4 provides the necessary background information on several HDR quality assessment metrics, subjective evaluation techniques and an overview of previous work done on HDR quality metric evaluation, tone-mapping operator evaluations and finally evaluation of HDR video compression algorithms.

- Chapter 5 describes the work done on evaluation of HDR videos over LDR videos from a viewers' perspective.

- Chapter 6 gives a detailed description of the evaluation of existing HDR video compression algorithms.

- Chapter 7 gives a detailed overview of the new HDR video compression algorithm which includes a novel analytical transfer function and an error minimisation scheme to encode the dynamic range and colour information, respectively.
- Chapter 8 concludes this thesis and presents possible future work.
- Appendix A, provides a detailed overview of the extensive HDR video quality evaluation framework which was created to objectively evaluate multiple HDR video compression algorithms at different quality levels against a number of HDR video sequences using a number video codecs and objective QA metrics.
- Finally, Appendix B provides a tone-mapped thumbnail representation of the 39 HDR video sequences used for the objective evaluations conducted in Chapter 6 and Chapter 7, respectively.

Chapter 2

Background

THE ultimate goal of imaging is to faithfully or artistically reproduce a scene as perceived by the HVS. The key difference between the HVS and existing imaging technology is the ability to perceive the entire range of lighting in a scene which cannot be reproduced by traditional digital imaging techniques, which, for the most part is limited by the storage format.

This primary focus of this chapter is to introduce the underlying concepts of dynamic range, HDR pipeline and file formats used to store real world lighting information. Additionally, it introduces other key concepts such as alternative colour spaces, image appearance models and tone mapping techniques, used frequently to process HDR image and video data. The concepts introduced herein will be used in the following chapters.

2.1 High Dynamic Range Imaging

This section introduces the fundamentals of dynamic range, the difference between Low/Standard Dynamic Range (LDR/SDR) and High Dynamic Range (HDR) imaging.

2.1.1 Light

In imaging systems, the physical measurement of light can be defined either by *luminance* or *spectral radiance*. Spectral radiance, in imaging systems, thus, can be defined as the radiant flux incident on a surface medium per unit wavelength, per unit solid angle (steradian), per unit projected area and can be mathematically defined by the derivative in equation 2.1:

$$L(\lambda) = \frac{d^2\phi(\lambda)}{d\omega.dA.\cos(\theta)} \tag{2.1}$$

where $L(\lambda)$ is the *spectral radiance* of the wavelength λ, ϕ is the radiant flux incident on the surface medium per unit time, ω is the solid angle, A is the area of the surface and θ is the angle between the incident ray and the surface [RHD*10].

Luminance, on the other hand, is the photometric measurement of luminous intensity per unit area of light travelling in a given direction and incident on the photo-receptor medium with a given solid angle and is measured in cd/m^2 physical units. Alternatively, it can also be defined as *spectral radiance*, integrated over all visible wavelengths as shown in equation 2.2.

$$Y = \int_{380nm}^{780nm} L(\lambda)V(\lambda)d\lambda \tag{2.2}$$

where $Y, L(\lambda)$ and $V(\lambda)$ represents the derived luminance, spectral radiance and a weighing function, respectively [Man06].

2.1.2 Dynamic Range

Dynamic Range (DR) is usually defined as the ratio between the brightest and the darkest luminance values present in a real-world scene or in an image. In photography and more generically in optics, DR of a scene is expressed in terms of *f-number*, sometimes also known as *f-stop* or *relative aperture* which is defined by the ratio between the lens's focal length and the opening diameter of the *relative aperture*. It is a dimensionless quantity expressed as $N = \frac{f}{D}$, where f is the focal length of the lens and D is the opening diameter of the relative (effective) aperture. The sequence of such numbers result in the change of luminance entering the photo-receptor by a power of 2 and is known as *f-stops*. The *f-stops* form a geometric series of powers of $\sqrt{2}$ such that $D \in [1, 1.4, 2, 2.8, 4, 5.6, 8, 11, 16, 22)$ and so on.

Alternatively, in digital imaging systems, the DR of a scene is defined as a ratio between the peak signal of the photo-receptor medium (maximum sensor capacity) and the noise level. Such a ratio is measured in decibels (dB) and can be mathematically defined as in equation 2.3.

$$DR_{scene} = 20log_{10}\frac{M_{sig.}}{RMS_{noise}} \tag{2.3}$$

where $M_{sig.}$ is peak signal capacity of the sensor and RMS_{noise} is the root-mean-square of the noise level. However, it should be noted here, that digital imaging systems are typically constrained by their storage format and bit depth. Therefore, in case the brightest absolute (physical) luminance value of a scene is greater than the sensor capacity, the same will be clamped to the maximum value of the storage format.

Furthermore, DR can also be defined as the contrast ratio between the brightest and the darkest luminance values of a scene, image or display and it can be stated as:

$$CR = \frac{L_{max}}{L_{min}} : 1 \tag{2.4}$$

where CR is the contrast ratio, L_{max} and L_{min} are the brightest and darkest luminance values of the scene/image/display.

Assimilating the three above mentioned representations of DR, they can however

be related and interchangeably used. For instance, a scene with a DR of 14 *f-stops* (as described in photography) has a contrast ratio of $2^{14} : 1 \equiv 16384 : 1$ which can be also be derived in terms of signal to noise ratio (SNR) as shown in equation 2.5.

$$\begin{aligned} f-stops &= log_2(10^{(\frac{dB}{10})}) \\ 14 &= log_2(10^{(\frac{dB}{10})}) \\ dB &= 10 \cdot log_{10}(2^{14}) \\ &= 42 \end{aligned} \tag{2.5}$$

Another example would be the contrast ratio of LCD displays. If assumed that an LCD monitor is perfectly black when all the pixels are set to zero, then the luminance of the screen is almost $0\ \mathrm{cd/m^2}$, therefore the DR is infinitely high. In reality, the minimum luminance of a good quality LCD monitor is around $1\ \mathrm{cd/m^2}$. Assuming the typical peak luminance value of a high quality LCD monitor is around $350\ \mathrm{cd/m^2}$, the contrast ratio is $350 : 1$.

2.1.3 Low Dynamic Range vs High Dynamic Range imaging

The concept of HDR is not new to photography. The best of the low-speed film stocks can offer up to 12 *f-stops* [Man06]. However, as mentioned earlier in Chapter 1, a majority of existing digital imaging systems can capture and display upto three orders of magnitude of DR i.e. a contrast ratio $CR = 1000 : 1$. In photographic terms, this translates to ≈ 10 *f-stops* or $\approx 30[\mathrm{dB}]$ in SNR terms. This is still less than what can be captured by dedicated HDR cameras or multi-exposure techniques (see Section 2.2.1 for further details). Therefore, the existing systems are typically termed as Low Dynamic Range (LDR) or Standard Dynamic Range (SDR) imaging systems and the limitations are largely due to the storage format used and is further discussed in Section 2.2.2.

On the other hand, from a purely optical perspective, although the light scattering effect can reduce the luminance and contrast perceived by the HVS to $\approx$ 2-3 orders of magnitude, the HVS is highly adaptive and can rapidly change the gaze and locally adapt to perceive luminance contrast of greater than four orders of magnitude. This, in turn, translates to an image or a scene with at least 14 *f-stops* and can thus be defined as an HDR image/scene [RHD*10]. Thus, in contrast to an LDR image, an HDR image has a contrast ratio of $\geq 16384 : 1$ or ≥ 42 [dB] in SNR terms [RHD*10].

2.2 High Dynamic Range pipeline

This section introduces the HDR pipeline which provides a brief end to end overview of the infrastructure and techniques required to capture, store, process and display HDR image and video content [BADC11]. The pipeline can be classified into four different stages namely:

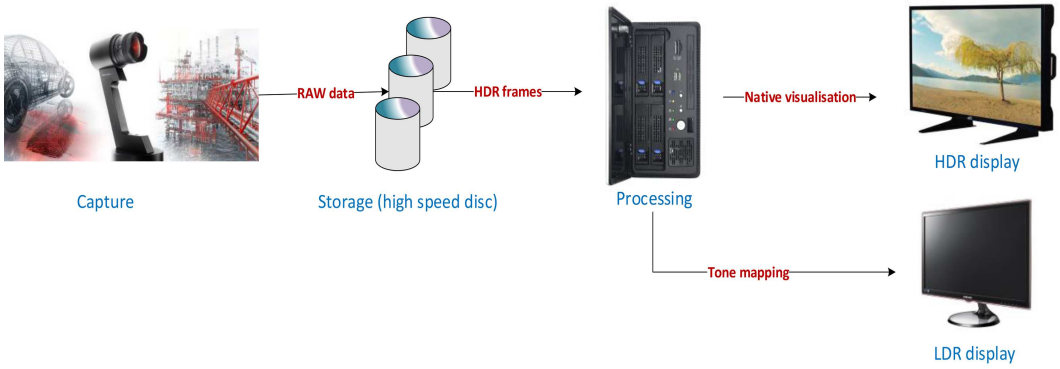


Figure 2.1: Schematic diagram of the generic HDR pipeline

acquisition, storage, processing and display. Sections 2.2.1, 2.2.2, 2.2.3 and 2.2.4 provide a brief overview of each stage of the pipeline.

2.2.1 Acquisition and content generation

Most consumer and professional level cameras are able to capture single exposure images in either traditional 8-bit JPEG (Joint Photographers Experts Group) format or are limited to 12/14 bit RAW format (.cr2 for Canon, .nef for Nikon, .pef for Pentax, .sr2 for Sony, .raw for Fuji etc.) which, although are based on the *.tiff* library, fail to cover the full range of lighting (from shadow to highlight details) present in most real world environments in a single exposure. Therefore, the most common way to generate HDR data using any medium to high-end digital camera, is by combining multiple bracketed exposures at different exposure settings to form an HDR image covering from lowest to highest luminance levels of the captured scene as proposed by Mann and Picard [MP94]. Assuming that the camera has a linear response [1], the radiance values stored in each exposure for each colour channel can be combined to recover the irradiance E as given in equation 2.6.

$$E(x) = \frac{\sum_{i=1}^{N_{exp}} \frac{1}{\Delta t_i} w(I_i(x)) I_i(x)}{\sum_{i=1}^{N} w(I_i(x))} \tag{2.6}$$

where I_i is the image captured at the i^{th} exposure, Δt_i is the exposure time of the image I_i, N_{exp} are the number of exposures and $w(I_i(x))$ is the weighting function which removes the outliers. In order to cover the entire range of lighting, the camera is set to capture the same scene at various levels of exposure values (Ev) with at least three exposures per scene. However, it is to be noted that higher number of overlapping exposures captured per scene

[1] in reality, the response function of most imaging systems follow an S-shaped curve which tends to saturate at the lowest and the highest luminance values. The middle portion has a response similar to a power or a logarithmic function. This non-linear compression is not to be confused with gamma correction [Man06].

ensures a better tonal gradation in the resultant HDR image [RHD*10, BADC11]. Figure 2.2 provides a visual description of HDR content generation using multi-exposure capture technique.

Figure 2.2: Multi-exposure technique to build HDR from single LDR exposures.

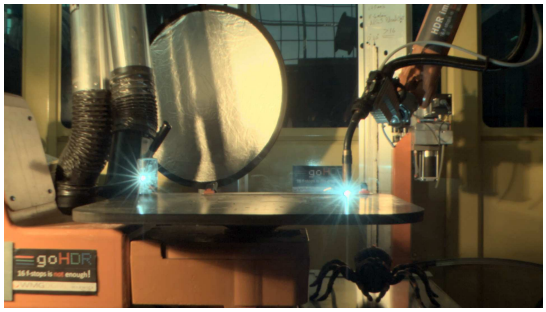
Another acquisition technique is to capture HDR static images natively using prototype HDR cameras such as the Spherocam HDR [AGb], the Panoscan MK-3 [Pan] and the Civetta 360 [AG"d]. These are high resolution 360-degree cameras which capture static HDR panoramas. Also, for HDR video capture, a number of prototypes have been built such as the HDRC CamCube [ic] and Spheron HDRv [AG"c]. These are multi-sensor imaging systems which capture HDR video with a native resolution of 640×480 and 1920×1080 at a standard 24 or 30 frames/sec (fps). Other commercially available professional cinema cameras such as the RED Epic [Com], ARRI Alexa [AG"a] and Black Magic Design [Des] can capture real-world lighting up to 14 f-stops. Figures 2.3a, 2.3b and 2.3c provides a tone-mapped representation of static HDR images/video frames captured using the Spherocam, HDRC Camcube and Spheron HDRv, respectively. Further details about HDR capturing techniques can be found in [RHD*10, BADC11] and [DLCMM16a].

Another alternative technique to generate HDR images is by expanding legacy LDR content by means of expansion operators, also known as Inverse Tone Mapping algorithms [BADC11]. These expansion operators endeavour to recreate HDR content applying techniques such as *blind inverse gamma function* and *radiometric calibration from a single image*. A detailed overview of these techniques are available in [BLDC06, RTS*07] and [BADC11].

Furthermore, HDR scenes can be synthetically created using computer graphics methods including ray tracing and rasterisation to render virtual scenes composed of for-

(a) Tone-mapped representation of an HDR image captured by the Spherocam HDR.

(b) Frame from an HDRC Camcube sequence (resolution 640×480).



(c) Frame from a Spheron HDRv sequence (resolution 1920×1080).

Figure 2.3: Tone-mapped representations of native HDR capture techniques.

mally defined geometric objects, materials and lighting, all captured from the perspective of a virtual camera. Similar to Figure 2.3a, a tone-mapped example of artificially generated HDR image is given in Figure 2.4.

Figure 2.4: Tone-mapped representation of a synthetically generated HDR image.

2.2.2 Data encoding and storage

In order to represent real-world luminance and chroma information with as much precision as possible, captured HDR data is typically post processed and stored as a 3-channel (RGB) floating point matrix [RHD*10]. However, the additional information results in large output files. For instance, an HDR image file with a full HD (1920×1080) resolution, represented by three single precision floating point colour channels would occupy approximately 24 MB of storage space (12 bytes per pixel). Therefore, it is evident that floating point formats are not entirely practical for distribution and transmission purposes [BADC11].

In order to mitigate this issue, several lossless HDR image/frame encoding formats have been proposed to date such as the Radiance RGBE (.hdr) [War91], OpenEXR (.exr) [FK13], LogLuv [Lar98] and its derivative, the Adaptive LogLuv [MT10]. Each of these file formats are described in brief detail later in Section 2.3.

2.2.3 Data compression

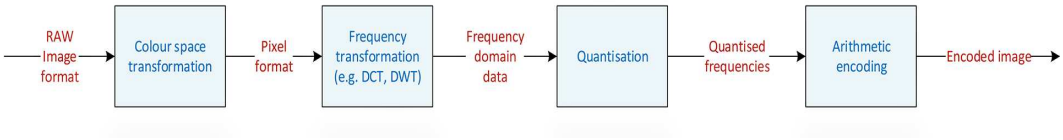


Figure 2.5: Lossy HDR image compression pipeline.

File formats discussed later in Section 2.3 are lossless techniques to encode real-world lighting information. Therefore, they are unable to provide effective compression to an extent feasible for mass storage and transmission purposes. To that extent, a number of lossy HDR image compression algorithms (techniques) have been proposed to date. These techniques can effectively be called *pre/post processing* techniques since they do not handle the final encoding which is typically performed by the JPEG/JPEG2000 engines. Figure 2.5 provides a visual description of the generic pipeline for lossy HDR image compression.

The earliest example of a pre/post processing technique was proposed by [WS04] by means of a sub-band encoding. *JPEG-HDR* [WS06], an extension of the previous proposal, provides a *backward-compatible* extension of the widely popular JPEG format to encode HDR images. This was later modified [WJN*12] to support additional features such as encoding of Wide Colour Gamut (WCG) content. Another lossy encoding format, the HDR-JPEG2000, proposed by Xu et al. [XPH05] extends to the popular JPEG2000 image format to encode HDR images by mapping the pixel values to a logarithmic domain in order to reduce coding errors. The mapped pixel values are subsequently encoded using the JPEG2000 encoder. Further details regarding the lossy compression of HDR images using the JPEG2000 format is available in [SK09]

Similar to lossy HDR image compression algorithms, several *pre/post processing*

algorithms for HDR video content have been proposed to date. The quintessential goal of such algorithms is to convert captured HDR frames [2] into a codec suitable format such that they can subsequently be encoded using available video codecs to produce an HDR video stream. Since, these algorithms and their subsequent evaluation form the core of this thesis, they are discussed in intricate detail, later in Chapter 3.

2.2.4 Display

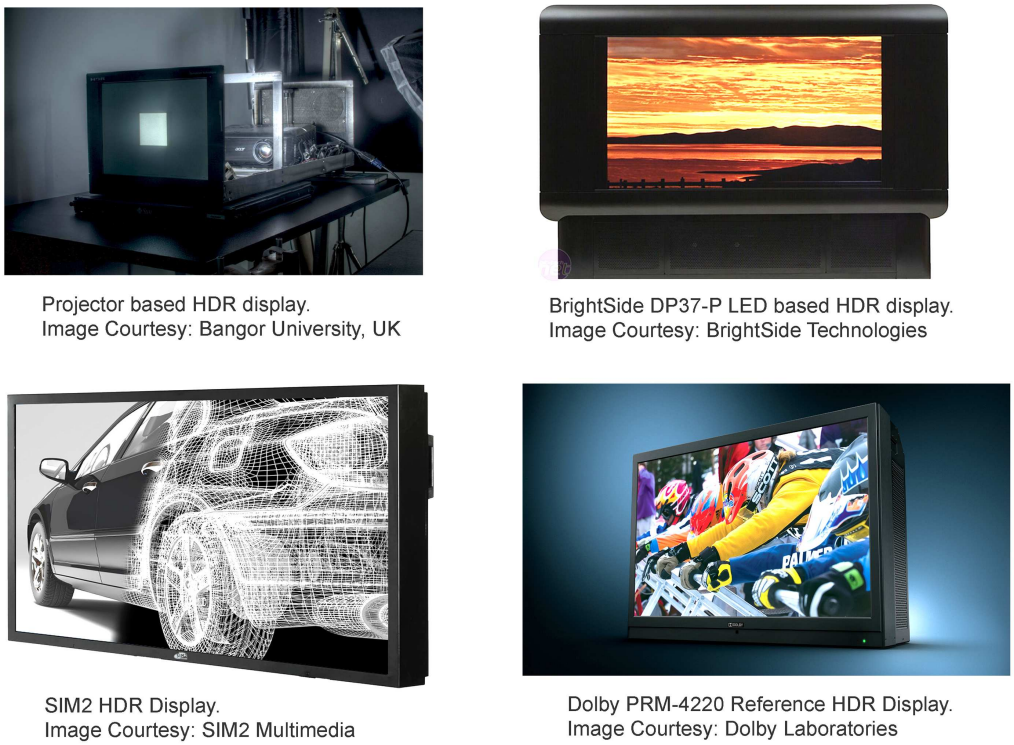


Figure 2.6: Native HDR display systems.

One of the primary goals of HDR imaging is to facilitate a more immersive experience to the viewer by providing a more accurate representation of the real-world lighting conditions. To that extent, the capabilities of the display technology have a major impact on the perceived image quality since the displayed HDR image/video is eventually processed by the HVS.

Typical conventional displays such as the Cathode Ray Tubes (CRTs) and Liquid Crystal Displays have a peak luminance of $\approx 80-200$ cd/m^2 with a limited contrast ratio [Des11, HR84]. Although, high-end Liquid Emission Displays (LEDs) and In-plane switching (IPS) [KP99] technologies can provide a slightly improved contrast ratio, they are still unable to provide the assumed immersive experience. In order to mitigate the issue,

[2] HDR video frames are stored in either .hdr or .exr format.

two different solutions of viewing HDR content have been proposed to date. HDR content can either be viewed in a native HDR display [SHS*04] or can be tone mapped, as discussed in Section 2.5 and displayed on a normal LDR monitor. The purpose of this section is to introduce the native HDR displays.

Display devices which support HDR video content such as the Brightside DR-37P (Dolby) [SHS*04, Sel13] and SIM2 [Sel13, SIMa] (see Figure 2.6) reference monitors rely on a technology termed as *dual-modulation displays*. The fundamental concept behind this approach is to optically combine the two display devices such that their intensities multiply to provide a very high intensity display device. The *dual-modulation displays* can be classified into two types such as the *Projector based display* and *LED based display* as explained below:

Projector based display

Seetzen et al. [SHS*04] developed the first prototype HDR display (Figure 2.6 - top left) based on *dual-modulation* technology. The basic modification introduced in this prototype display was to insert a second light modulator in the form a projector (providing uniform back light) behind a conventional LCD panel and coupling the two light modulators using a Fresnel lens and diffuser thus increasing the brightness of the back light.

The assembly comprised of three components namely,

- A conventional 15" XGA colour LCD panel driven by an analogue-to-digital LCD controller allowing VGA connection. (1^{st} *modulator*)
- A digital light processing (DLP) projector (Optoma DLP EzPro 737) with its colour wheel removed, resulting a monochrome projector with a threefold increase in light intensity (2^{nd} *modulator*).
- A projection lens for the projector and the Fresnel lens to collimate the projected backlight into a narrow viewing angle in order to increase maximum brightness and avoid colour distortion. Finally, an LCD diffuser was used to redistribute the collimated light into a reasonable viewing angle. (*Optics to couple the modulators*)

The three components were installed in a single housing and aligned to create a close match between the DLP and LCD pixels. The final prototype achieved a contrast ratio of 50000:1 with a peak luminance of 2700 cd/m^2 and a minimum luminance of 0.54 cd/m^2. However, projector based HDR displays suffer from several major drawbacks. Apart from the obvious form-factor due to a long optical path required by the projector, power consumption, cost factor, thermal management and video bandwidth are some of the major drawbacks of projector based HDR displays. Nonetheless, these types of displays were able to prove that native viewing of HDR content was possible even though unsuitable for commercial production.

LED based display

Seetzen *et al.* (2004) developed a second LED based HDR display (Figure .2.6 - top right). The authors took advantage of the concept known as *veiling glare*, which impacts the visibility of details in dark areas next to very bright regions due to the light scattering effect inside the human eye. Instead of a projector, this display had an LCD panel as the 1^{st} modulator responsible for details and colour and a low resolution LED panel as the 2^{nd} modulator which provided additional back light. The square root of the luminance in the image was down sampled to match the resolution of the LED panel and approximately de-convoluted by the LED point spread function (PSF) [SHS*04]. The prototype DR 37P consisted of 1380 LEDs and a 37" LCD panel capable of producing a contrast ratio of $200,000:1$ with a peak luminance value measured at 3000 cd/m^2 and a minimum luminance of 0.015 cd/m^2 [Sel13].

The use of LED overcomes the power consumption and heat issues. Form-factor is no longer an issue as the LED back light panel are only as thick as conventional LCD back light. The video bandwidth requirements are reduced due to the low resolution of the LED panel. Therefore, an LED powered HDR display removes the commercialisation barriers which the previous projector powered display suffer from. The first commercial HDR display, the *Solar 47* (Figure 2.6 - bottom left) was released by SIM2 in 2009 with a full HD resolution and 2206 LEDs and supports processing 16 bits/channel images.

2.3 HDR file formats

The HDR content generated needs to be stored, distributed and post-processed. The floating point HDR pixel values are represented using 3-channel single precision floating point matrices, assuming three channels for primary (RGB) colours. However, images represented by single precision floating point can take up to 96 bpp. This results in a file much larger than its equivalent LDR image format (.jpg/.png) without compression. Therefore, efficient and compact representations were required to store HDR content and address the high memory demands. As mentioned previously in Section 2.2.2, the three most popular HDR content encoding file-formats are RGBE (.hdr), OpenEXR (.exr) and LogLuv (.tiff). This section describes each of these formats in detail.

2.3.1 Radiance RGBE (.hdr) format

Ward [War91] proposed the first solution to this problem. The Radiance (RGBE) format was introduced as part of the *Radiance Lighting Simulation and Rendering System* [War94b] and is widely used for HDR photography and image-based lighting. The pixel data is either encoded as 32 bit RGBE encoding or its CIE variant XYZE encoding with 3-bytes for colour components R_m, G_m and B_m and 1-byte for the exponent E. The component E is

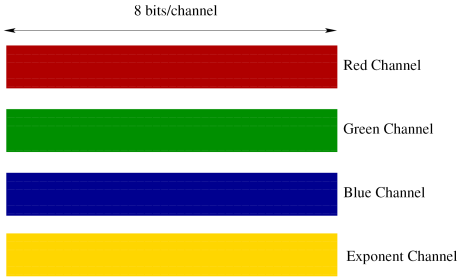


Figure 2.7: Bit breakdown for the RGBE 32-bit integer representation

share between the three colour channels, assuming that it does not vary much. The RGBE format can encode a dynamic range of up to 76 orders of magnitude and can be formulated as:

$$E = \lceil \log_2(max(R_w, G_w, B_w)) + 128 \rceil \tag{2.7a}$$

$$R_m = \lfloor \frac{256 R_w}{2^{E-128}} \rfloor \tag{2.7b}$$

$$G_m = \lfloor \frac{256 G_w}{2^{E-128}} \rfloor \tag{2.7c}$$

$$B_m = \lfloor \frac{256 B_w}{2^{E-128}} \rfloor \tag{2.7d}$$

and the decoding format as:

$$R_w = \frac{R_m + 0.5}{256} 2^{E-128} \tag{2.7e}$$

$$G_w = \frac{G_m + 0.5}{256} 2^{E-128} \tag{2.7f}$$

$$B_w = \frac{B_m + 0.5}{256} 2^{E-128} \tag{2.7g}$$

2.3.2 OpenEXR RGBA (.exr) format

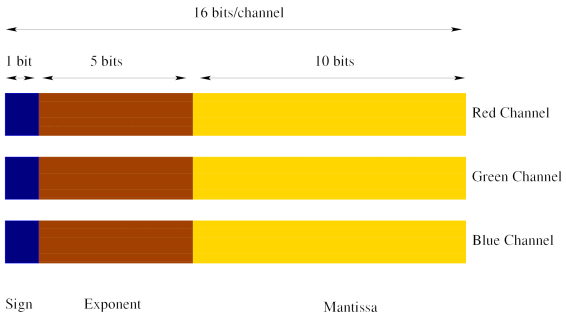


Figure 2.8: Bit breakdown for the OpenEXR Half Pixel encoding

The Extended Range format was made available as an open source C++ library in

2002 by Industrial Light & Magic. It is based on a 16-bit half floating point type. Each RGB pixel occupies 48 bits, each channel broken into 16-bit floating point format with 1-sign , 5-exponent and 10-mantissa bits. The .exr format covers up to 10.7 orders of magnitude with a relative error less that 5% and quantization step size less that 0.1%. The OpenEXR library also supports 32 bits/channel (96 bpp) and 24 bits/channel (72 bpp) float data type introduced by Pixar. However, the most widely used is the 16 bits/channel half-float representation. Further details can be obtained in [FK13].

2.3.3 LogLuv (.tiff) format

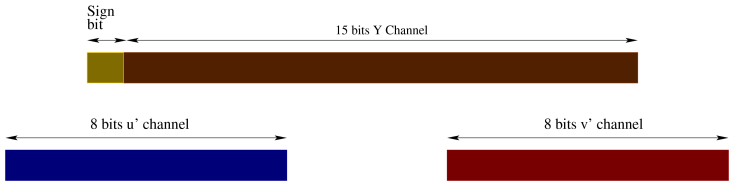


Figure 2.9: Bit breakdown for the LogLUV (.tiff) encoding format

The LogLuv encoding [Lar98] was introduced as a perception based colour encoding for scene-referred images. Similar to the IEEE floating point RGB encoding (sometimes referred to as the ultimate HDR image representation offering up to 96bpp), LogLuv was also implemented as part of the TIFF library. There are two variants of LogLuv colour encoding namely, the 24-bit LogLuv and the 32-bit LogLuv encoding. For most practical purposes, the 32-bit version is used. Therefore, in this discussion, only the 32-bit version is explained. Modelled on the HVS which does not perceive a real world scene as strongly co-related RGB but as *brightness* and *colour* separately, the encoding segregates luminance and chroma allocating 16-bits for luminance in the logarithmic domain and 8-bits for each of the two chroma channels in a linear domain thus covering 38 orders of magnitude in 0.3% steps. The two chroma channels u' and v' are calculated according to perceptually uniform CIE chromaticity scales. Further details about the CIE uniform chromaticity scales are explained later in Section 2.4. The 32 bpp format encoding can be formulated as:

$$L_{15} = \lfloor 256(\log_2 Y_w + 64) \rfloor, \quad u_8 = \lfloor 410u' \rfloor \quad ,v_8 = \lfloor 410v' \rfloor \tag{2.8}$$

where L_{15}, u_8 and v_8 are the integer representation of the luminance and chroma channels, respectively. Similarly, the decoding can be formulated as

$$Y_w = 2^{\frac{L_{15}+0.5}{256}-64} \quad u' = \lfloor \frac{u_8}{410} \rfloor \quad ,v' = \lfloor \frac{v_8}{410} \rfloor \tag{2.9}$$

where Y_w, u' and v' are the floating point representations of the luminance and chroma channels, respectively.

2.3.4 Adaptive LogLuv format

Motra & Thoma [MT10] proposed a modified and optimised version of the LogLuv transform for HDR image/video frame encoding. The authors argue that the LogLuv transformation from single precision floating point data to 16-bit integer representation of luminance is suboptimal in terms of bit-depth allocation especially for video encoding since state-of-the-art video encoders can support up to 14-bits/channel [AMT]. Furthermore, most scenes do not utilise the full dynamic range. Therefore, the modification facilitates transformation of real-world luminance Y to 'n'-bit luma [3] L between the representable range $[Y_{min}, Y_{max}]$ of the image/frame. To convert HDR frames (stored in RGB colour space) to $Lu'v'$ format, the frames are first converted from RGB to $Yu'v'$ colour space as mentioned later in Section 2.4. Subsequently, the luminance channel Y is mapped to 'n' -bit luma using the following system of equations:

$$L = \begin{cases} 0 & \text{if } Y = 0 \\ \lfloor a(\log_2 Y + b) & \text{otherwise} \end{cases}$$

and the inverse mapping from 'n'-bit luma to real-world luminance is formulated as:

$$Y = \begin{cases} 0 & \text{if } L = 0 \\ 2^{(\frac{L+0.5}{a} - b)} & \text{otherwise} \end{cases}$$

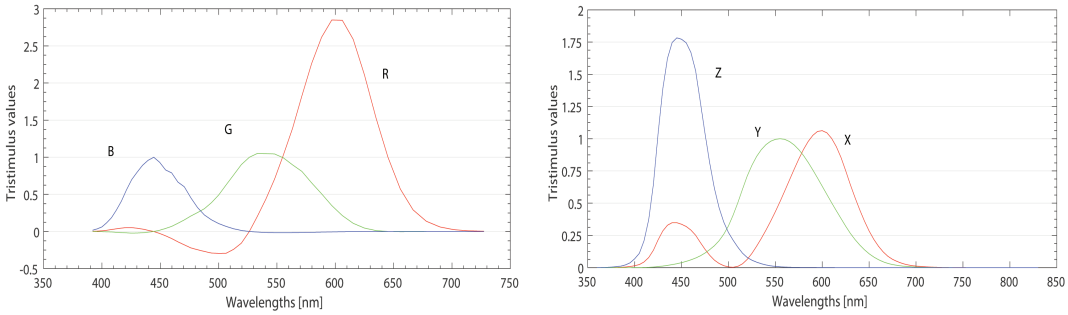
where $a = \frac{2^n - 2}{\log_2 \frac{max(Y)}{min\_positive(Y)}}$ and $b = \frac{1}{a} - \log_2(min\_positive(Y))$.

Note that if Adaptive LogLuv transform is used, the parameters a and b need to be passed along with the $Lu'v'$ image/video frame such that they can be used to inverse map the values. However, conversion from RGB to LogLuv is a lossy process because the transform quantizes the RGB data into perceptual bins. The transformation parameters a and b can be adapted on a per image/frame basis or on slices (part of the image/frame) or on a group of pictures (GOP) basis.

2.4 Colour

Light can be defined precisely by measuring physical units of spectral radiance but the human perception of colour is a totally different entity. Colourimetry is the science which links human colour perception and the physical description of light. This section introduces the fundamental aspects of colour spaces which is used throughout the thesis. A detailed introduction can be found in [RKAJ08, Fai13a], while [WS82] and [Hun05a] contain more

[3] Originally coined by NTSC to differentiate between physical luminance and video signal brightness, luma can be defined as $L = f(Y)$, where f is a transfer function and Y is the physical luminance. Further details about mapping from luminance to luma are given in Section 3.2.



(a) The CIE 10-degree observer RGB colour matching functions.

(b) XYZ colour matching functions.

Figure 2.10: Plot of 10-degree observer colour matching functions. Image courtesy *www.cvrl.org*.

comprehensive information about colourimetry.

Human colour perception is defined by three types of cones sensitive to Long ($\approx$ L $\in (560, 580)$ nm), Medium ($\approx$ M $\in (530, 540)$ nm) and Short ($\approx$ S $\in [420, 440]$ nm) wavelengths. Numerically, light is a multi-dimensional variable wherein each dimension is associated with a specific wavelength [Man06]. The colour visible to the human eye, is a mapping of this multi-dimensional variable to three primaries corresponding to three types of cones. The projection can be mathematically defined as a product spectral power distribution $\phi(\lambda)$ and the spectral response of the type of cones $C_L(\lambda), C_M(\lambda)$ and $C_S(\lambda)$ [Man06]. The product is mapped to the three primaries [4] R, G and B as:

$$R = \int_{\lambda} \phi(\lambda) C_L(\lambda) d\lambda \tag{2.10a}$$

$$G = \int_{\lambda} \phi(\lambda) C_M(\lambda) d\lambda \tag{2.10b}$$

$$B = \int_{\lambda} \phi(\lambda) C_S(\lambda) d\lambda \tag{2.10c}$$

$$\tag{2.10d}$$

Now, due to the 3D encoding of colour, the number of uniquely distinguishable colours to the HVS is limited. Additionally, for each spectral target, the intensity of the primary stimuli are adjusted to visibly match the target stimulus. Subsequently, for each spectral target stimulus with a specific wavelength λ, the adjusted intensities of the primary stimuli can be modelled by three functions such as $r(\lambda), g(\lambda)$ and $b(\lambda)$, also known as "colour matching functions". Therefore, from the colour matching experiments, the target stimulus can be visibly matched by the linear combination of the primary stimuli and is

[4] Colour matching experiments [WS82] conclude that virtually all colours visible to the HVS can be matched by adding light from three visibly pure stimuli also known as the "primary stimuli", which are Red (R), Green (G) and Blue (B).

mathematically defined as:

$$Q_\lambda = r(\lambda)\cdot R + g(\lambda)\cdot G + b(\lambda)\cdot B \tag{2.11}$$

where R, G and B are scalar multipliers. Since the primaries have a fixed wavelength i.e. $\lambda_R = 645.2nm, \lambda_G = 525.3nm$ and $\lambda_B = 444.4nm$ [SB59], the target stimulus is represented as a linear combination of the triplet (R, G, B) and are known as the tristimulus value of the target Q_λ. However, the colour matching functions, as shown in Figure 2.10a, resulted in negative values of the R primary to represent colours which are too saturated to be within the visible range. Therefore, for mathematical convenience (since it is easier to deal to with colour spaces where the tristimulus values are always positive), CIE defined [HP85] defined an alternative set of colour matching functions where any Q_λ can be matched with positive primary coefficients. The colour matching functions are known as $x(\lambda), y(\lambda)$ and $z(\lambda)$ and a 2D plot of the functions are given in Figure 2.10b. Similar to equation 2.11, Q_λ can be formulated as:

$$Q_\lambda = x(\lambda)\cdot X + y(\lambda)\cdot Y + z(\lambda)\cdot Z \tag{2.12}$$

where X, Y and Z are the tristimulus values.

Alternatively, for a given target stimulus Q_λ, the tristimulus (X, Y, Z) values can be obtained as:

$$X = \int_{380}^{830} Q_\lambda x(\lambda) d\lambda \tag{2.13a}$$

$$Y = \int_{380}^{830} Q_\lambda y(\lambda) d\lambda \tag{2.13b}$$

$$Z = \int_{380}^{830} Q_\lambda z(\lambda) d\lambda \tag{2.13c}$$

For the sake of convenience, colours are often represented in a 2D space using chromaticity coordinates which can be derived from the XYZ tristimulus values as in equation 2.14

$$\begin{aligned} x &= \frac{X}{X+Y+Z} \\ y &= \frac{Y}{X+Y+Z} \end{aligned} \tag{2.14}$$

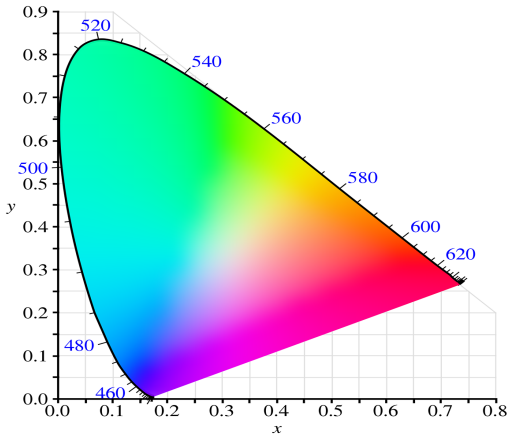
In addition, the coordinate z can also be formulated as $z = \dfrac{Z}{X+Y+Z} = 1-x-y$. However, as z can be defined in terms and x and y, only the latter is accompanied by the luminance Y to describe the colour. This two-dimensional CIE xy chromaticity space can be plotted in a diagram as shown in Figure 2.11a, which defines the number of distinguishable colours that the HVS can perceive. However, the visible differences between colours are not very well defined by the xy chromaticity coordinates. Therefore, in 1976, CIE defined the

uniform chromaticity scales (UCS), u' and v' to uniquely define the perceptual differences between colours and they are formulated as in equation 2.15.

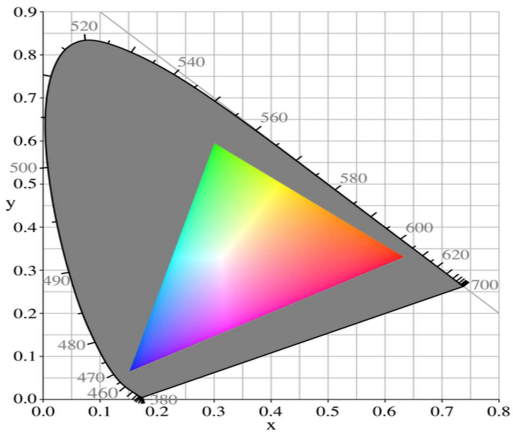
$$u' = \frac{4X}{X + 15Y + 3Z} \tag{2.15a}$$

$$v' = \frac{9Y}{X + 15Y + 3Z} \tag{2.15b}$$

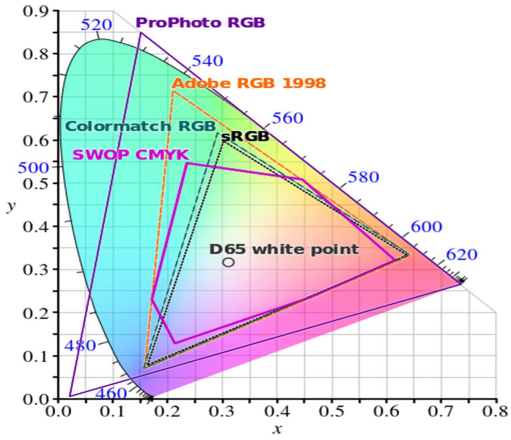
2.4.1 Colour gamut



(a) The CIE 1931 2D chromaticity diagram.



(b) Typical sRGB (REC. 709) gamut displayed by LDR monitors.



(c) Comparison of RGB colour gamuts.

Figure 2.11: Plot of CIE xy chromaticity coordinates along with different RGB gamuts.

As described previously, any given colour can be realised/matched using a set of three given primaries and the values of those primaries can be mapped to any three specific points in the chromaticity diagram thus forming a triangle. The area under the triangle

contains the range of colours that are realisable given the three primaries and is known as the *"colour gamut"*. Colours which cannot be represented given a set of primaries and maps to an area outside the triangle are known as *"out of gamut"* colours.

In figure 2.11b, the triangle exhibits the range of colours realisable using the primaries defined by the ITU-R Recommendation BT.709 (also known as the BT.709/REC.709 primaries) [Int02]. The gamut boundary marks the approximate range of colours which can be displayed using a standard CRT/LCD monitor and is also known as the sRGB colour space. Although, sRGB is the most widely used colour space in modern digital imaging applications, several other wider colour gamuts exist for specific applications where the defined primaries ensure a wider range of realisable colours. Figure 2.11c exhibits the areas under the several colour gamuts used in modern digital applications.

2.4.2 Colour spaces and white points

A colour space is a specific organisation of colours inside a 2D boundary defined by a colour vector (R,G,B triplet) which forms the colour gamut. It can also be defined as the relationship between the colour gamut and the standard CIE XYZ colour space. Alternatively, it can also be defined as a mathematical model which describes the way colours can be quantified using the colour vector.

Since colour spaces are typically a set of formulas which define the relationship between a given colour vector and the standard CIE XYZ space [RHD*10], the transformation from one colour space to another is generally linear, often performed using a 3×3 colour transformation matrix, albeit with a few exceptions where the relationship is non-linear and additional formulas are required for the transformation [RHD*10]. Before introducing colour space transformations, a brief overview of white point is required since linear or non-linear transformations are dependent on the assumed white point.

A white point of a display (it is display specific) is the colour emitted by the display when all three constituent channels (R, G and B) contribute equally. An often used white point illuminant is the CIE D_{65} illuminant which is assumed to the reference light source in case there are no further information available about the white point or the illuminant of the scene being captured or displayed. Further information about white point illuminants is available in [RHD*10, RKAJ08].

Now, in order to convert from one tristimulus colour space to another, the xy chromaticity coordinates of the primaries i.e. $(\mathrm{x_R}, \mathrm{y_R}), (\mathrm{x_G}, \mathrm{y_G})$ and $(\mathrm{x_B}, \mathrm{y_B})$ needs to be known. In addition, the white point in xy chromaticity coordinates $(\mathrm{x_W}, \mathrm{y_W})$ and the peak luminance value Y_W needs to be specified.

Given the xy chromaticity coordinates of the primaries as well as the white point, the z coordinates (z_R, z_G, z_B and z_W) can be computed. Subsequently, from the peak luminance Y_W and the chromaticity coordinates of the white point, the tristimulus values of the white point $(\mathrm{X_W}, \mathrm{Y_W}, \mathrm{Z_W})$ can be computed using the inverse of the formula given in equation

2.14. Now, using the known white point tristimulus values, a system of linear equations can be solved to determine the *conversion constants* to be used for converting any given (R,G,B) colour vector triplet to the CIE XYZ colour space as shown in equation 2.16.

$$\begin{aligned} X_W &= x_R C_R + x_G C_G + x_B C_B \\ Y_W &= y_R C_R + y_G C_G + y_B C_B \\ Z_W &= z_R C_R + z_G C_G + z_B C_B \end{aligned} \tag{2.16}$$

where C_R, C_G and C_B are the constants. Finally, using the computed constants a generic 3×3 conversion matrix can be formulated to convert the RGB colour space to CIE XYZ as shown in equation 2.17.

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} x_R C_R & x_G C_G & x_B C_B \\ y_R C_R & y_G C_G & y_B C_B \\ z_R C_R & z_G C_G & z_B C_B \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{2.17}$$

One of the most popular RGB to XYZ conversion matrix used in modern digital imaging and the one used throughout this thesis is the ITU-R recommendation BT.709 3×3 conversion matrix. Equations 2.18 and 2.19 provide the forward and inverse transformation matrices.

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.4124 & 0.3576 & 0.1805 \\ 0.2126 & 0.7152 & 0.0722 \\ 0.0193 & 0.1192 & 0.9505 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{2.18}$$

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 3.2405 & -1.5371 & -0.4985 \\ -0.9693 & 1.8760 & 0.0416 \\ 0.0556 & -0.2040 & 1.0572 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \tag{2.19}$$

Although the CIE-XYZ is considered to be the absolute colour space, values in the XYZ are not always realisable due to negative gamut values. Also, the pixel values are in an RGB image are highly correlated with each other resulting to undesired changes in correlated channels when pixel values of any one of the three channels are manipulated [RKAJ08]. Furthermore, the HVS is more sensitive to luminance (brightness) variations than chromatic variations. This feature has often been exploited by many image and video encoding algorithms where the RGB images are first converted to a colour space which de-correlates the luminance and chroma components for compression and processing purposes. Subsequently the chroma components are presented at a lower resolution (sub-sampled/ down-sampled) than the luminance component. One such popular colour space which is most frequently used for broadcasting purposes (HDTV standard) is the YC_bC_r colour space [IR] where Y is the luminance component and C_b and C_r are the chroma components.

From a compression perspective, RGB images/video frames are typically converted

to the YC_bC_r space using the transformation matrix presented in equation 2.20.

$$\begin{bmatrix} Y \\ C_b \\ C_r \end{bmatrix} = \begin{bmatrix} 0.299 & 0.5186 & 0.114 \\ -0.168 & -0.333 & 0.498 \\ 0.498 & -0.417 & -0.081 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{2.20}$$

This is followed by a sub-sampling of the chroma components as discussed later in Section 3.4.4. The compressed image/video frame is then up-sampled and an inverse conversion to RGB is performed using the transformation given in equation 2.21.

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 1.000 & 0.000 & 1.397 \\ 1.000 & -0.343 & -0.711 \\ 1.000 & 1.765 & 0.000 \end{bmatrix} \begin{bmatrix} Y \\ C_b \\ C_r \end{bmatrix} \tag{2.21}$$

However, for HDR specific applications, a modified version $Yu'v'$ is often used [Lar98] where u' and v' define the uniform chromaticity scales. Using the BT.709 primaries [Int02], the RGB values are converted to the XYZ colour space as defined in equation 2.18. Subsequently, the u' and v' values are computed as in equation 2.15 [5].

The fundamentals of the colour spaces and transformations discussed till now has been used throughout this thesis. However, any discussion on colour spaces is incomplete without a brief overview of perceptually uniform colour spaces. The next section provides a brief overview of perceptually uniform spaces including the state-of-the-art IPT colour opponent space used later in Chapter 7.

2.4.3 Perceptually uniform colour spaces

The two primary issues with RGB, CIE XYZ, YC_bC_r and $Yu'v'$ colour spaces is that they are not perceptually uniform and inter-channel correlations exist. Perceptual uniformity essentially means that the perceived difference between any two colours, say C_1 and C_2 is not equal to the euclidean distance between them. This can be explained as follows:

Let C_1 and C_2 be any two colour vector triplets (in any non-uniform space) where (p_1, q_1, r_1) and (p_2, q_2, r_2) be the vector values. Therefore, the euclidean distance ΔC between C_1 and C_2 can be formulated as:

$$\begin{aligned} \Delta C &= \sqrt{(\Delta p)^2 + (\Delta q)^2 + (\Delta r)^2} \\ \therefore \Delta C &= \sqrt{(p_1 - p_2)^2 + (q_1 - q_2)^2 + (r_1 - r_2)^2} \end{aligned} \tag{2.22}$$

Since perceptual uniformity is advantageous for digital image manipulation purposes [RKAJ08], two perceptually uniform colour spaces, the CIE 1976 $L^*u^*v^*$ and CIE $L^*a^*b^*$, henceforth abbreviated as CIELUV and CIELAB were defined to provide perceptually uni-

[5] The luminance component can also be separately computed using the formula $\mathrm{Y} = 0.2126 \cdot \mathrm{R} + 0.7152 \cdot \mathrm{G} + 0.0722 \cdot \mathrm{B}$

formity. The derivations from CIE 1931 XYZ tristimulus values to CIELUV and CIELAB are shown in equations 2.23 and 2.24, respectively.

The CIELUV [RKAJ08] colour space is formulated as:

$$L^* = 116(\frac{Y}{Y_n})^{1/3} - 16 \tag{2.23a}$$

$$u^* = 13L^*(u' - u'_n) \tag{2.23b}$$

$$v^* = 13L^*(v' - v'_n) \tag{2.23c}$$

The conversion holds under the assumption that $\frac{Y}{Y_n} > 0.008856$. If $\frac{Y}{Y_n} < 0.008856$ then L^*_m is defined as: $L^*_m = 903.3\frac{Y}{Y_n}$ and u'_n, v'_n are derived as:

$$u'_n = \frac{4X_n}{X_n + 15Y_n + 3Z_n} \tag{2.23d}$$

$$v'_n = \frac{9Y_n}{X_n + 15Y_n + 3Z_n} \tag{2.23e}$$

The CIELAB [RKAJ08] follows a similar approach. For each of the ratios $\frac{X}{X_n}, \frac{Y}{Y_n}$ and $\frac{Z}{Z_n} > 0.008856$, the space is defined as:

$$L^* = 116(\frac{Y}{Y_n})^{1/3} - 16 \tag{2.24a}$$

$$a^* = 500[(\frac{X}{X_n})^{1/3} - (\frac{Y}{Y_n})^{1/3}] \tag{2.24b}$$

$$b^* = 200[(\frac{Y}{Y_n})^{1/3} - (\frac{Z}{Z_n})^{1/3}] \tag{2.24c}$$

and for ratios > 0.008856, the colour space is defined as:

$$L^*_m = 903.3\frac{Y}{Y_n} \tag{2.24d}$$

$$a^*_m = 500[f\frac{X}{X_n} - f\frac{Y}{Y_n}] \tag{2.24e}$$

$$b^*_m = 200[f\frac{Y}{Y_n} - f\frac{Z}{Z_n}] \tag{2.24f}$$

where the function $f(.)$ is defined as

$$f(r) = \begin{cases} r^{\frac{1}{3}} & \text{for } r \geq 0.008856 \\ 7.787r + \frac{16}{116} & \text{for } r \leq 0.008856 \end{cases}$$

Although CIELUV and CIELAB provide perceptual uniformity to a good degree, based on the latest psychophysical data, it has been assessed that some non-uniformities do remain [RHD*10]. Also CIELUV and CIELAB are prone to hue compression issues [RKAJ08].

To mitigate the issues, a more recent perceptually uniform space was proposed in the form of the IPT colour space.

The IPT colour space [EF98a] was designed to improve upon CIELUV and CIELAB with respect to hue uniformity. Similar to CIELUV, the luminance component is the I channel which stands for intensity. The chroma components are the P & T channels which stands for Protan and Tritan, respectively. The following transformations are required to convert RGB pixel values to the IPT space.

First, the RGB values are converted to XYZ tristimulus values using equation 2.18. Second, the XYZ values are converted to the LMS cone excitation space using the Hunt-Pointer-Estevez transformation matrix as shown in equation 2.25. The matrix defines cone responses under D_{65} white point illumination.

$$\begin{bmatrix} L \\ M \\ S \end{bmatrix} = \begin{bmatrix} 0.4002 & 0.7075 & -0.0807 \\ -0.2280 & 1.1500 & 0.0612 \\ 0.0000 & 0.0000 & 0.9184 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \tag{2.25}$$

The linearly transformed LMS cone responses are then non-linearly transformed for spectral sharpening [RKAJ08].

$$L' = \begin{cases} L^{0.43} & \text{if } L \geq 0 \\ -(-L)^{0.43} & \text{if } L < 0 \end{cases}$$

$$M' = \begin{cases} M^{0.43} & \text{if } M \geq 0 \\ -(-M)^{0.43} & \text{if } M < 0 \end{cases}$$

$$S' = \begin{cases} S^{0.43} & \text{if } M \geq 0 \\ -(-S)^{0.43} & \text{if } M < 0 \end{cases}$$

Finally, the spectrally sharpened $L'M'S'$ values are transformed to the IPT space using equation 2.26

$$\begin{bmatrix} I \\ P \\ T \end{bmatrix} = \begin{bmatrix} 0.4000 & 0.4000 & 0.2000 \\ 4.4550 & -4.8510 & 0.3960 \\ 0.8056 & 0.3572 & -1.1628 \end{bmatrix} \begin{bmatrix} L' \\ M' \\ S' \end{bmatrix} \tag{2.26}$$

The backward transformations from IPT to RGB can correspondingly be computed by inverting the matrices defined in equations 2.26 and 2.25 and linearising the spectral sharpening. Further details are beyond the scope of this thesis and extensive details of colour spaces has been covered previously in [RKAJ08].

2.5 Tone Mapping

Tone mapping can be defined as the process of mapping source static images/video sequences with high contrast-high dynamic range and wide colour gamut to a destination medium with limited contrast and colour reproduction capabilities [MMS99]. Typically, this refers to mapping HDR images/video sequences with scene-referred pixel values representing physical (absolute) luminance to device referred (relative) pixel values in order the match the capabilities of traditional 8-bit LDR display systems. The process can be formulated as:

$$f(I) : Src_i^{whc} \rightarrow Dest_o^{whc} \tag{2.27}$$

where Src_i and $Dest_o$ are the source input and destination output, respectively. w, h and c are the width, height and number of colour channels of the input and output image/frame, respectively.

Algorithms, which perform the mentioned mapping are termed as Tone Mapping Operators (TMOs). Previous literature has classified the TMOs on the basis of the mathematical operation(s) and design philosophy. While the mathematical classification leads to the grouping of TMOs as either global, local, frequency based or segmentation operators, TMO's grouped based on design philosophy can be classified as perceptual or temporal TMOs. A detailed overview of each of these categories along with the TMOs which can be grouped into each of these categories is given in [BADC11]. Furthermore, TMOs can also be classified in terms of their intent such as Visual System Simulators (VSS), Scene Reproduction Operators (SRO) and Best Subjective Quality operators (BSQ) [EWMU13]. However, it should be noted that the large volume of available TMOs make it impossible to cover all aspects of tone mapping [6] within these classifications and there are exceptional cases where certain operators serve special requirements [Man06].

This section provides a brief discussion on some of the widely popular TMOs based on their mathematical classification, some of which have been used in the work outlined later in Chapters 5 and 6.

2.5.1 Global TMOs

Global TMOs preserve global contrast because all pixels are equally treated. The operator may sometimes perform a first pass of the image to calculate image statistics, which are subsequently used to optimize the dynamic range reduction. The simplicity of global TMOs facilitates their extention into a temporally coherent TMO by introducing motion information and compensation. However, the major disadvantage of global TMOs is the loss of fine details and local contrast due to strong quantization. Popular global TMOs proposed are [Sch95a, War94a, LRP97, PFFG98, DMAC03] and [RSSF02].

[6] thousands of papers have been published on TMOs and evaluation of TMOs. As of May 2016, a google scholar search with exact title "tone-mapping" reveals 8600 results.

2.5.2 Local TMOs

Local TMOs on the other hand, preserve local image contrast thereby improving the appearance of the tone-mapped image. Instead of performing the tone-mapping on the whole image, the operator(s) subdivides the image in small 8×8 blocks and applies f based on the neighbourhood of each pixel. The final image appears better compared to global TMOs but are more prone to artefacts such as prominent halos around edges. Due to their inherent design, local TMOs are computationally more expensive and are much more difficult to extend into the temporal domain by introducing motion information and compensation. Popular local TMOs proposed are [CHS*93, Ash02, LSC04] and [RSSF02].

2.5.3 Frequency based TMOs

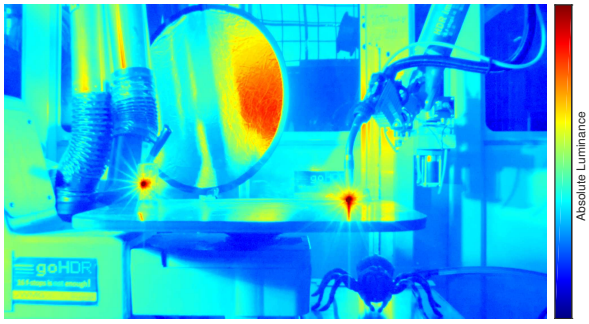
Frequency based TMOs share the same goal as the local TMOs i.e. to preserve edges and local contrast. However, unlike the global and local TMOs, the mathematical operations are performed in the frequency domain instead of the usual spatial domain. However, the limitations of these operators are that they only achieve better results that local TMOs if and only if a total separation between the large details and edges is achieved [BADC11]. These operators can further be divided by their approach to the tone mapping problem. Such division include Low Curvature Image Simplification [TT99], Bilateral Filtering [TM98, PD09] and Gradient Domain Tone Compression [FLW02].

2.5.4 Segmentation based TMOs

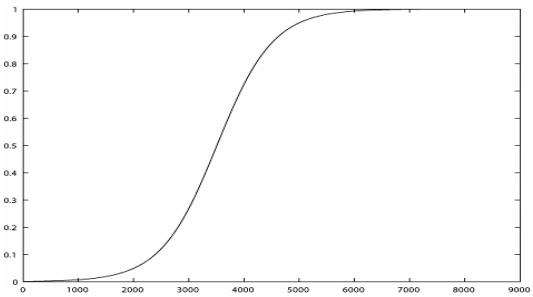
Segmentation based TMOs rely on the divide-and-conquer approach to the tone mapping problem. Source HDR images are divided into uniform segments and a global mapping function is applied on each segment. The divided segments are subsequently merged to form the complete tone-mapped image/frame. This approach can be stated as a compromise between a purely global and purely local operator and the primary advantage of such an approach is that gamut modifications are minimised as linear operators suffice in many cases. A few widely used TMOs [YP03, KMS05, LFUS06, MKVR09] follow this approach. Further details are available in [BADC11].

2.5.5 The Photographic Tone Reproduction Operator (Reinhard TMO)

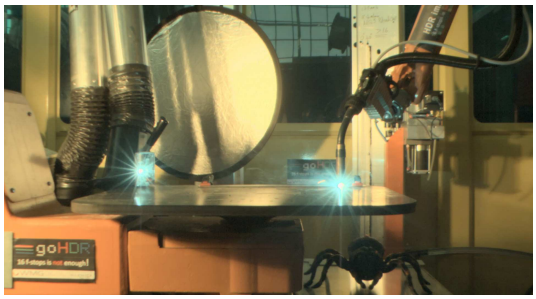
Amongst the hundreds of TMOs available for various tone-mapping purposes, one of the most widely used TMO is the Photographic Tone Reproduction Operator, proposed by Reinhard et al. [RSSF02] which was inspired by a zone metering system [Ada80, Ada81, Ada83], originally proposed by Ansel Adams. The motivation behind specifically mentioning this operator is primarily due to its repeated usage for HDR video compression purposes by existing compression algorithms. Since HDR video compression forms the core of this thesis, a brief overview of this operator is provided herein.



(a) False Colour representation of HDR



(b) Sigmoidal tone compression



(c) Tone-mapped representation using Reinhard global TMO

Figure 2.12: Sigmoidal compression of the Photographic TMO

The Reinhard TMO simulates the dodging and burning effect that photographers have applied to captured films for more than a century. Such a simulation results in an *S-shaped* tone compression curve which closely follows the perceptual attributes of the HVS. A simplified version of this operator can be defined as:

$$L_d(x,y) = \frac{L_{(x,y)}}{1+L_{(x,y)}} \tag{2.28}$$

where L_d is the displayable luminance, $L_{(x,y)}$ is the physical luminance scaled by a factor of αL_w^{-1}. α is the chosen exposure (in film analogy) and L_w is the logarithmic average of the scene. The denominator causes a graceful blend between these two scalings and the formulation which replicates *sigmoidal* tone compression is guaranteed to bring all luminance within displayable range $L_d \in [0,1]$ (which can be discretised to an 8-bit integer range, $L_d \in [0,255]$, as shown in Figure 2.12b). However, this effect is not desirable in high-contrast scenes. Therefore, equation 2.28 can be extended allowing high luminance values to burn out in a more controllable fashion. To that extent, equation 2.28 was combined with linear mapping, resulting in the *global photographic tone reproduction operator*

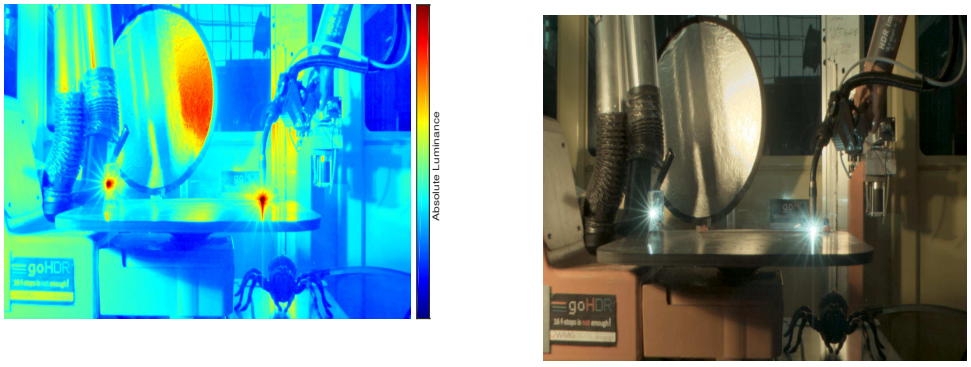
given in equation 2.29

$$L_d(x,y) = \frac{L(x,y)(1+\frac{L(x,y)}{L_{white}^2})}{1+L(x,y)} \tag{2.29}$$

where L_{white} is the smallest luminance values which can be mapped to white (or burnt out in photographic terms).

The local version was, essentially, the global operator applied for smaller image regions typically 8×8 neighbourhoods. The local operator is less computationally efficient than the global operator. However, the performance can be improved by executing a scale selection mechanism (neighbourhood scale) on the fly. Furthermore, the photographic tone reproduction operator has also been extended to facilitate temporal coherence for video tone-mapping purposes [KRTT12]. Figure 2.12 provides a visual description of the photographic TMO.

2.5.6 Display Adaptive Tone Mapping (Mantiuk)



(a) False Colour representation of HDR

(b) Tone-mapped representation using Display Adaptive TMO

Figure 2.13: False Colour representation and equivalent tone-mapped representation using Display Adaptive TMO

Mantiuk et al. [MDK08] proposed a TMO where the primary goal is to preserve the appearance of the original HDR scene including contrast, sharpness and colours by adjusting the image/video content with the pre-notion of the ambient illumination and capabilities of the target display. Such a TMO is also known as a Scene Reproduction Operator (SRO) [EWMU13]. The TMO produces the least distorted image in terms of visible contrast distortions given the characteristics of the target display. In order to produce the least distorted image, the distortions are weighted with an HVS model which takes into account luminance masking, spatial contrast sensitivity and contrast masking.

The authors demonstrate that such a TMO can be defined as a non-linear optimisation problem where the error function is weighted by the HVS model and the limitations of the target display controls the constraints of the output image. This non-linear optimisation

problem can be simplified by reducing the degrees of freedom of the optimised system. The reduction in the degrees of freedom facilitates an efficient solution where the problem can be reduced to a quadratic equation. Furthermore, a straightforward extension of this TMO to account for temporal coherence allows it to be used for HDR video sequence tone-mapping. By reducing the degrees of freedom of the optimised system, the authors introduced a TMO with adjustable parameters that employs a piecewise linear tone-curve to map the HDR luminance to its corresponding LDR luminance. The authors state although high-contrast images require local tone-mapping to retain details, a well designed piecewise tone-curve can produce good results and provide maximum flexibility. However, this TMO does not take into account the colour appearance issues, as the authors were unable to find a robust colour appearance model. Instead, the desaturated colour-to-luminance ratios introduced by Schlick [Sch95b] was used to preserve the chroma information as shown in equation 2.30.

$$R' = \left(\frac{R}{L}\right)^{s} L' \tag{2.30}$$

where L is the luminance, R is the trichromatic value, L' is the tone-mapped luma and R' is the tone-mapped colour channel. The work assumes the saturation value of $s = 0.6$.

Based on the study conducted by Yoshida et al. [YBMS05a], it was found that consumers preferred sharper images which was attained by using contrast enhancement techniques which would increase the contrast of the displayed image by upto 100%. However, to avoid over-processing, the image enhancement techniques used in this work enhances the contrast of the *reference* image by 15% [Hun05b]. The TMO uses a display model, as shown in equation 2.31 to account for the limitations of the target display and an HVS model based on Daly's contrast sensitivity function (CSF) [Dal92] to derive a piecewise tone-curve which maps the *reference* HDR luminance to a Just Noticeable Difference (JND) space such that the visible distortions due to the luminance mapping are minimised.

$$L_d(L) = (L')^{\gamma} \cdot (L_{max} - L_{black}) + L_{black} + L_{reflect} \tag{2.31}$$

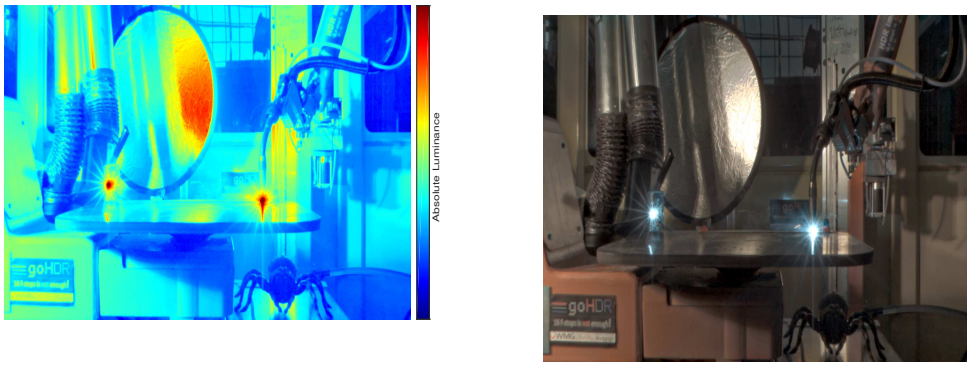
where L_d is the displayed luminance, L' is the nomalised pixel value $pix_{value} \in (0,1]$, $\gamma \approx 2.2$, L_{max} is the peak display luminance, L_{black} is the display black level and $L_{reflect}$ is the ambient light that is reflected from the surface of the target display. It is defined as:

$$L_{reflect} = \frac{k}{\pi} E_{ambience} \tag{2.32}$$

where $E_{ambience}$ is the ambient luminance in lux and k is the reflectivity of the display surface. The tone-mapped frame were subsequently compared with the original 0^{th} (base) LDR image by means of subjective evaluation (pairwise-comparison) and results demonstrate that the tone-mapped image was preferred over the base exposure for both dark and bright environments. An overview of the HVS model used in this TMO is given later in

section 3.2 and further details of the TMO is available in [MDK08].

2.5.7 iCAM06 - Image Appearance Model (iCAM)



(a) False Colour representation of HDR

(b) Tone-mapped representation using Display Adaptive TMO

Figure 2.14: False Colour representation and equivalent tone-mapped representation using Display Adaptive TMO

An image appearance model is an extension of a colour appearance model [Fai13b] incorporating properties of spatial and temporal vision thus allowing prediction of complex input stimuli. Such a model endeavours to predict the perceptual response complex spatial stimuli, thereby also providing a scope of predicting the appearance of an HDR image. Kuang et al. [KJF07] proposed a new image appearance model, designated iCAM06, designed specifically for HDR image rendering. Based on the iCAM framework [MFH*02], the new model incorporates the spatial processing models in the HVS for contrast enhancement and photo-receptor light adaptation functions that enhance local details in highlights and shadows. The new model inherits several modules from the original iCAM model which includes local white point, chromatic adaptation and usage of perceptually uniform colour space such as IPT (see Section 2.4.2). However, it also includes several improvements for enhanced prediction of HDR renderings and production of more aesthetically pleasing images.

The original HDR image is first converted to the CIE-XYZ colour space using the CIE 1931 XYZ tristimulus values and subsequently decomposed into a base and a detail layer wherein the base layer is obtained using an edge-preserving bilateral filter [PD09] and the detail layer is obtained by subtracting the base layer from the original image.

The base layer first undergoes chromatic adaptation which is achieved by converting the pixel values in XYZ colour space to a spectrally sharpened RGB image using the M_{CAT02} transformation matrix [MFH*02] as shown in equation 2.33. The incomplete adaptation factor is computed as a function of adaptation luminance and the surround factor as shown in equation 2.34 where L_A is computed as 20% of adaptation white and F is the surround

factor. Finally, the adaptation factor is used in the chromatic adaptation transformation of the tri-stimulus values as shown in equation 2.35.

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 0.7328 & 0.4296 & -0.1624 \\ -0.7036 & 1.6975 & 0.0061 \\ 0.0030 & 0.0136 & 0.9834 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \tag{2.33}$$

$$D = 0.3F \left[1 - \frac{1}{3.6} e^{\left(\frac{-(L_A - 42)}{92} \right)} \right] \tag{2.34}$$

$$\begin{aligned} R_c &= \left[\left(R_{D65} \frac{D}{R_w} \right) (1 - D) \right] R \\ G_c &= \left[\left(G_{D65} \frac{D}{G_w} \right) (1 - D) \right] G \\ B_c &= \left[\left(B_{D65} \frac{D}{B_w} \right) (1 - D) \right] B \end{aligned} \tag{2.35}$$

The chromatic adaptation also converts the global white point to CIE illuminant D65 [NR05] which is used by the IPT transformation at a later stage.

Subsequently, the spectrally sharpened RGB image is converted from the CAT02 space (M^{-1}_{CAT02}) to the Hunt-Pointer-Estevez fundamentals (M_{HPE}) which is where the resultant RGB signal undergoes a non-linear tone compression using a response function for both rods and cones derived from the Hunt Model [Hun91] as shown in equation 2.36.

$$\begin{bmatrix} R' \\ G' \\ B' \end{bmatrix} = M^{-1}_{CAT02} \cdot M_{HPE} \begin{bmatrix} R_c \\ G_c \\ B_c \end{bmatrix} \tag{2.36}$$

The tone-compressed R'G'B' image is then converted to the perceptually uniform colour opponent space IPT [EF98a], which is desired for image attribute adjustments without affecting other attributes. To preserve the naturalness of the rendered tone-compressed image, the detail layer is enhanced to predict the Stevens effect and the P & T channels of the base layer is enhanced to predict the Hunt effect [Fai13b]. Finally, the enhanced base and detail layers are combined to produce an enhanced perceptually uniform output image. This is displayed on the target device by converting the IPT image to an RGB signal followed by an inverse chromatic adaptation.

The authors demonstrate, by means of a pairwise-comparison based subjective evaluation that the proposed image appearance model performed significantly better than four other TMOs compared in this work. Further details are available in [KJF07].

2.6 Summary

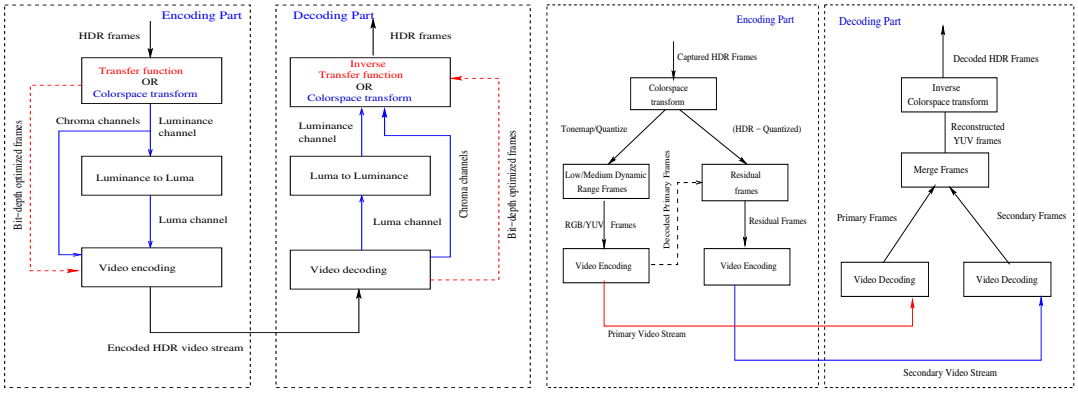
This chapter provides a brief overview of the critcial underlying concepts of HDR imaging. Sections 2.1 and 2.2 provides an overview of HDR imaging pipeline including a detailed discussion on capture, storage, processing and display mechanisms. These underlying concepts are used throughout the thesis. Section 2.4 provides a brief overview of the critical colour space transformations required for efficient manipulation of HDR image/video data. These transformations are critical in tone-mapping and compression applications as outlined later in Chapters 5, 6 and 7. Finally, although not the central topic of this thesis, Section 2.5 provides a brief overview of tone-mapping and discusses some of the TMOs which have been used later in Chapters 5 and 6.

Chapter 3

High Dynamic Range Video Compression

THIS chapter introduces the reader to the fundamental concepts of HDR video encoding. First, the two generic approaches to HDR video content compression are introduced to the reader. This is followed by a detailed discussion on each of the two approaches and the state-of-the-art video compression algorithms which follow either of the approaches. The chapter also discusses the fundamental concepts of transfer functions and their usage in HDR video compression. Finally, an overview of state-of-the-art video codecs [1] used for compression purposes is hereby provided.

3.1 Generic approaches to HDR video compression



(a) Non-backward compatible HDR video compression (b) Backward compatible compression

Figure 3.1: Schematic diagram of the two generic approaches to HDR video compression

[1] A codec or a *coder-decoder* is a software for encoding and decoding of digital data stream. Further details about codecs is given later in Section 3.4.

Legacy video codecs which are responsible in encoding video data into usable (viewable) formats are only able to support integer data up to 14 bits/pixel/channel such that the input data range $\in [0, 16383]$. Typically, legacy videos are captured in 8-bit formats. Also, a sequences of images, captured in legacy 8-bits/pixel/channel formats such as .jpg and/or .png can be directly used by native video codecs to encode into resultant video streams. However, as discussed previously in Chapter 2 - Section 2.2.2, HDR video content captured in *.hdr/.exr* floating point format poses major storage and transmission issues since they cannot be directly encoded by native video codecs.

HDR video compression algorithms which form the core of this thesis are quintessentially *pre/post processing* algorithms whose primary goal is to convert floating point HDR data to a video codec suitable intermediate file format, typically *.yuv* or *.y4m* which can be used by native video codecs to encode the HDR information into a video stream as shown by the schematic (black box) diagram in Figure 3.1 [MDBR*16a]. These lossy pre/post processing algorithms can be classified into two approaches i.e. the *non-backward* compatible approach and the *backward* compatible approach. Sections 3.1.1 and 3.1.2 provide an overview of the two approaches, respectively. Further details about individual compression algorithms which follow either of the two approaches is given in Section 3.3.

3.1.1 Non-backward compatible approach

The *non-backward* approach to HDR video compression, as shown in Figure 3.1a takes advantage of the higher bit-depth encoding support (typically 10-14 bits) in state-of-the-art video codecs such as H.264/AVC [WSBL03, AMT] and HEVC [SOHW12] in order to pack as much lighting and colour information as possible into a single HDR video stream. This is typically achieved using a range of transfer functions (perceptual and opto-electronic) which convert floating point information to an $M-bit$ integer format where M is typically 10-14 bits/pixel/channel. On the decoder side, the compressed video stream can be decoded to an HDR display or tone-mapped onto an LDR display. However, the primary limitation of this approach is that existing hardware (FPGA[2]) based video decoders are unable to decode higher bit-depth encoded videos. Thus, the compressed videos can only be played back using customised hardware or software-based video decoders.

3.1.2 Backward compatible approach

Since LDR video file formats encoded using legacy codecs such as MPEG2 have become widely supported by all software and hardware equipment, it is impractical to expect an immediate replacement with their HDR counterparts. To facilitate a smooth transition from LDR to HDR, the *backward* compatible approach as shown in Figure 3.1b,

[2]A field-programmable gate array is an integrated circuit designed to be configured for specific task(s) and contains large resources of logic gates and RAM blocks.

splits the input HDR video stream into a base and a residual stream, each with a fixed number of bits, typically 8-10. The resultant video streams can thus be encoded and decoded using any legacy codec and decoder, respectively. A few *backward compatible* algorithms [WS06, MEMS06, LK08] contain an 8-bit/channel (24 bits/pixel) tone-mapped (TM) stream which enables it to be played on legacy video players. The residual stream contains additional information to be used for reconstruction of HDR frames during decompression. These algorithms typically use *dual-loop* encoding where the base stream is first encoded and decoded back to create the residual stream. The residual stream compensates for the distortions introduced by the codec at a chosen compression quality level, thus minimising the loss in quality during the reconstruction of HDR frames.

3.2 Transfer functions

This section introduces the fundamental concepts of transfer functions (TFs) as used in HDR image and video compression. Ideally, the concept of TFs form a key part of tone-mapping and several TMOs are essentially TFs. However, the concept was not introduced earlier since its relevance in the context of this thesis is limited to HDR video compression only.

A TF is ideally a mathematically reversible function $f(\cdot)$ such that $f(\cdot) : y \longrightarrow L$, where y represents the range of input pixel values of the frame in absolute/physical luminance terms such that $y \in [10^{-5}, 10^{9}]\ cd/m^2$. L, represents the range of output code/luma values which is typically dependent on the bit-depth support of the codec. Thus, if an *n-bit* codec is used to encode the HDR video frames then $L \in [0, 2^n - 1]$. For HDR video compression purposes, TFs are typically modelled on the HVS response to physical luminance. These models are fundamentally non-linear ranging from power [Ste57] to logarithmic response functions [WRM96] or a combination of both [MMS06, SYD87].

Although, various models have been suggested to date, the HVS is highly complicated and adaptive. Therefore, it is quite difficult to accurately model the HVS response to real-world luminance [MMS06]. In order to understand the basic HVS response to input luminance, it is to be noted that although the HVS cannot accurately determine the exact magnitude of absolute luminance, it can consistently detect contrast differences, also termed as luminance difference Δy from the background luminance y [SYD87]. The basic derivation of a TF can thus be formulated from the Weber's law [WRM96] where the contrast can be defined as in equation 3.1.

$$C = \frac{\Delta y}{y} \tag{3.1}$$

Now, if the minimum contrast (threshold contrast) C_t to detect differences is measured at various background luminance levels, it can be shown that the two asymptotic regions

are connected via a transition slope [SYD87] and can be referred to as the *Weber-Fechner* relationship where regions of zero slope follows the Weber's law as shown in equation 3.2.

$$C_t = \frac{\Delta y}{y} = k(constant). \tag{3.2}$$

The psychophysical experiments yielding the *Weber-Fechner* relationship involve the detection of a small varying target luminance whose luminance varies from a uniform background. Based on these experiments the value of constant C_t has been detected to as low as 0.01. Now, based on the *Weber-Fechner* relationship, the visual response of the HVS can be modelled as in equation 3.3

$$\begin{aligned} \Delta L &= C_t \frac{\Delta y}{y} \\ \implies dL &= C_t \frac{1}{y} dy \end{aligned} \tag{3.3}$$

Integrating over the whole range of physical luminance values implies

$$\begin{aligned} \int dL &= C_t \int \frac{1}{y} dy \\ \implies L &= C_t log(y) + C \end{aligned} \tag{3.4}$$

where C is an arbitrary constant. This basic logarithmic TF derived from the *Weber-Fechner* relationship can be applied to HDR video compression. However, in order to do so, certain assumptions need to be made. For instance, assuming that the range of physical luminance $y \in [10^{-5}, 10^{9}]$ and a 12-bit video codec [AMT] is used, $L \in [0, 2^{12} - 1]$, the task is to find a function which satisfies the following properties:

1. `Property 1:` The function can encode the full range of input physical luminance within the specified bit-depth of the codec.

2. `Property 2:` A unit distance in the output range L correlates with Just Noticeable Difference (JND) which offers more uniform distribution and thus more control over distortions which are inevitably introduced during lossy compression processes.

3. `Property 3:` Only positive integer values are to be used to encode L since this simplifies video compression.

4. `Property 4:` A half-unit distance in L is below one JND such that the loss is image quality during compression cannot be perceived by the HVS.

In order to derive such a function, let $t(y_{adapt})$ be a function which provides a conservative estimate of the detection threshold Δy against an adaptation luminance y_{adapt}. To satisfy property 3, it needs to be ensured that the rounding of code values of L to nearest integer does not introduce visible distortions which in turn also satisfies property 4 and

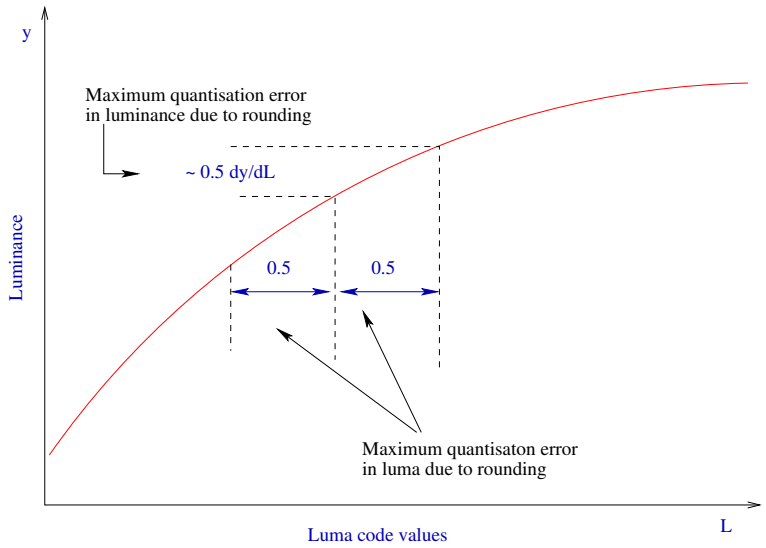


Figure 3.2: Quantisation error in luma code values expressed in terms of luminance such that error $<$ 1JND [MMS06].

property 2. The maximum allowable quantisation error due to rounding of code values L can be ± 0.5. Since, the detection thresholds are measured in terms of a physical luminance, the luma quantisation errors need to converted to luminance as shown in Figure 3.2. This can be done by the Taylor series expansion of the function $y(L)$. Therefore, the quantisation errors can be redefined as:

$$y(L+0.5) - y(L) \approx 0.5\frac{dy}{dL} \tag{3.5}$$

To satisfy property 4, the function $t(y_{adapt})$ needs to satisfy the inequality shown in equation 3.6.

$$\frac{1}{2}\frac{dy}{dL} < t(y_{adapt}) \tag{3.6}$$

The above inequality can be rewritten as:

$$\frac{dy}{dL} = 2 \cdot \frac{t(y_{adapt})}{k} \tag{3.7}$$

It is to be noted that k is at least ≥ 1 and larger values of k result in lower quantisation error albeit at the cost of more code values required to encode the entire dynamic range of y. Rewriting inequality 3.6 as equality 3.7 results in a form where a differential change in luma code values L results in a differential change in physical luminance $y(L)$. By assuming $y_{adapt} = y$, $C_t = 0.01$ (as determined by equation 3.2), equation 3.7 can be simplified by replacing $t(y_{adapt}) = C_t \cdot y_{adapt}$ which implies $t(y_{adapt}) = 0.01 \cdot y$. Now the simplified

equation 3.7 can be solved by taking an integral as shown in equation 3.8.

$$\begin{aligned} \int dL &= \frac{k}{2}\int \frac{1}{t(y_{adapt})}dy \\ \therefore \int dL &= \frac{k}{2}\int \frac{1}{0.01\cdot y}dy \\ \therefore \int dL &= 50k\int \frac{1}{y}dy \\ \implies L &= 50k\cdot log(y)+C \end{aligned} \tag{3.8}$$

where C is an arbitrary constant. Given the boundary value conditions i.e $y \in [10^{-5}, 10^{9}]$ and $L \in [0, 2^{12}-1]$, the values of k and c can be obtained by solving two simultaneous equations as shown below:

$$\begin{aligned} 4095 &= 50k\cdot log(10^{9})+C \\ 0 &= 50k\cdot log(10^{-5})+C \end{aligned} \tag{3.9}$$

which results in $k = 2.54$ and $C = 1463$ respectively. Using the obtained values of k and c, the relation between physical luminance y and luma code values can finally be written as:

$$\boxed{L = \lfloor 127\cdot log(y)+1463 \rfloor} \tag{3.10}$$

Therefore, it is clearly evident that equation 3.10 satisfies property 1 of the function $f(\cdot)$. Figure 3.3 provides the visual description of the logarithmic response function as derived. Similarly, it can be shown that by changing the boundary conditions for both y and L, the

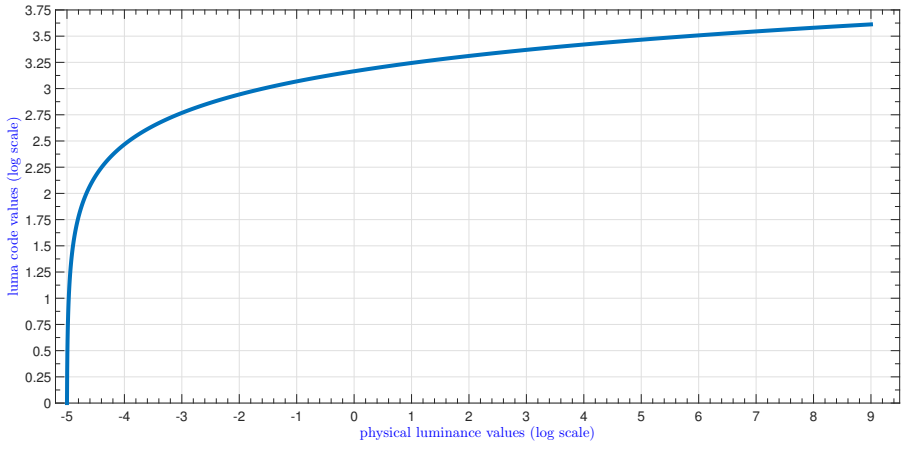


Figure 3.3: Luminance vs Luma – the logarithmic response function

logarithmic response functions can be derived for 10-, 12- and 14-bit codecs. Alternative derivations of the TF are given in [MMS06] and [BMP15]. However, as observed in

previous psychophysical experiments, the *Weber-Fechner* relationship only holds true for luminance levels $\geq 100cd/m^2$ [SYD87]. Consequently, the HVS response to luminance levels below $100cd/m^2$ can be modelled more precisely by the *De Vries-Rose* relationship which uses a $\frac{1}{2}$-power model of perceived brightness versus input luminance [Kel77] and is defined as:

$$L = C_1 y^{\frac{1}{2}} + C_2 \tag{3.11}$$

where C_1 and C_2 are arbitrary constants. Also, it has been shown that the HVS response for luminance levels $\leq 10cd/m^2$ is approximately linear [DVMS74]. Therefore, from the combined evidence available in [DVMS74, SYD87, MMS06], it is quite evident that accurate modelling of an HVS response function is a complex task and a logarithmic model is often unable to accurately predict the HVS response to input luminance. Furthermore, from the computational perspective a logarithmic response function is inefficient as precious code values (bits) are wasted to accurately map regions of lower luminance with redundant precision at the cost of lesser precision at regions of higher luminance [Man06]. Consequently, more precise HVS response functions were proposed for HDR image and video compression purposes based on further psychophysical experiments [MMS06]. The proposed TFs, based on their target requirements can classified into two groups namely Perceptual TFs (PTFs) and Opto-Electronic TFs (OETFs). Sections 3.2.1 and 3.2.2 provide a brief introduction to PTFs and OETFs, respectively.

3.2.1 Perceptual transfer function (PTF)

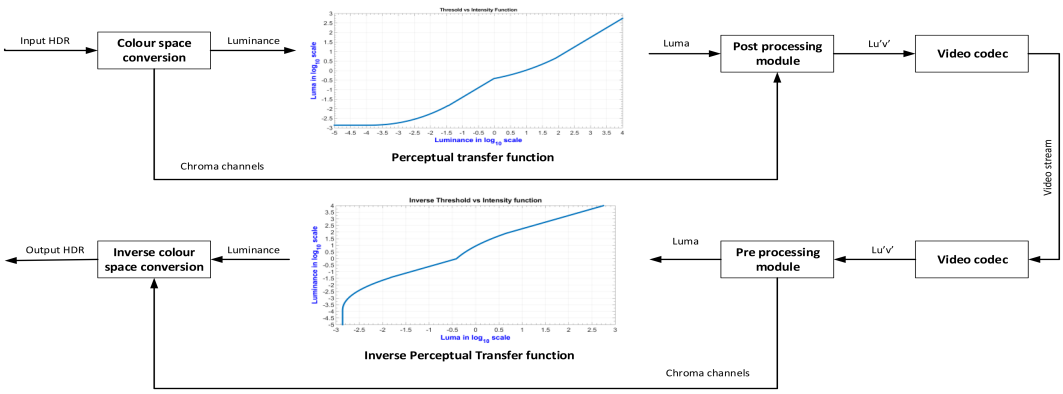


Figure 3.4: Generic schema of PTF based HDR video compression

A Perceptual Transfer Function (PTF) is a TF which endeavours to accurately model the photoreceptor response to physical luminance. Such a TF is generally derived from a *threshold versus intensity (tvi)* function which predicts the minimum difference of the target luminance against a background adaptation luminance y_{adapt}. The *tvi* function is derived from a Contrast Sensitivity function (CSF) which describes the loss of sensitivity of the HVS as a function of spatial frequency and adaptation luminance. A detailed description of

CSFs can be found in [Bar92, Bar99, Bar03, SYD87, Man06, MKRH11].

For HDR image and video compression purposes, it is redundant to model the HVS response with absolute accuracy. Therefore, a more simplistic approach as used by [Man06] is to convert luminance values to a non-linear space (luma values) that is scaled in JND units. Such a space guarantees that addition and/or subtraction of '1' in this space results in a just noticeable difference in perceivable brightness. If $f(\cdot)$ is a function which converts the JND scaled luma L to physical luminance y then this can be written as $f : L \longrightarrow y$. Using the definition of a *tvi* function, this can be re-written as:

$$f(L+1) - f(L) = tvi(y_{adapt}) \tag{3.12}$$

By using a Taylor series expansion:

$$f(L+1) = f(L) + \frac{df(L)}{dL} \tag{3.13}$$

the LHS of the equation 3.12 can be replaced and re-written with its first-order approximation:

$$\frac{df(L)}{dL} = tvi(y_{adapt}) \tag{3.14}$$

Here, an assumption that the HVS can adapt to a single pixel value of y is introduced to simplify the equation such that $y_{adapt} \equiv y = f(L)$ [SYD87]. Thus, equation 3.14 can be re-written and $f(L)$ can be obtained by solving the first order differential equation as shown in equation 3.15

$$\begin{aligned} \frac{df(L)}{dL} &= tvi(y) \\ \therefore \frac{df(L)}{dL} &= tvi(f(L)) \end{aligned} \tag{3.15}$$

Since *tvi* functions are also mathematically reversible, the inverse of $f(\cdot)$, i.e. $\psi = f^{-1}(\cdot)$ can be used to map physical luminance to a JND scaled luma such that:

$$\boxed{L = \psi(y)} \tag{3.16}$$

This function ψ as shown in the form of equation 3.16 is typically used for HDR image and video compression purposes. The actual shape of the *tvi* function has been extensively studied [Bar03, Dal92, FPSG96, VMV72, BB71] and found to be strictly monotonic.

Based on the CSF data,obtained from several psychophysical experiments, several *tvi* functions have been proposed to date. Barten et al. [Bar92] proposed the Grayscale Display Function (GDF) which was adopted by the DICOM standard [MDG08] for medical imaging purposes. The GDF was one of the first PTFs based on a CSF data [Bar92] validated through psychophysical experiments which maps a physical luminance range of

$y \in [0.05, 4000]$ to a 10-bit JND scaled luma space $L \in [0, 1023]$. Other psychophysically based *tvi* functions include the response function proposed by Ferwarda et al. [FPSG96], widely used for several computer graphics applications, the Meeteran CSF model [VMV72] which was improved by Kodak and used in the Visible Difference Predictor [Dal92], the global and local cone-response function [SYD87] and the *tvi* model suggested by Bodmann [Bod73] based on Blackwell's CSF data [BB71] which was adopted by the CIE standard [Bla81].

Although this section introduces the reader to the basic derivation of a PTF, the detailed usage of such PTFs for HDR image and video compression is shown later in Section 7.1.2 where several commonly used PTFs are used in the form of a plug-and-play structure as a part of a generic framework for *non-backward* compatible HDR video compression. A generic diagram of a PTF based HDR video compression scheme is given in Figure 3.5.

3.2.2 Opto-Electronic transfer function (OETF)

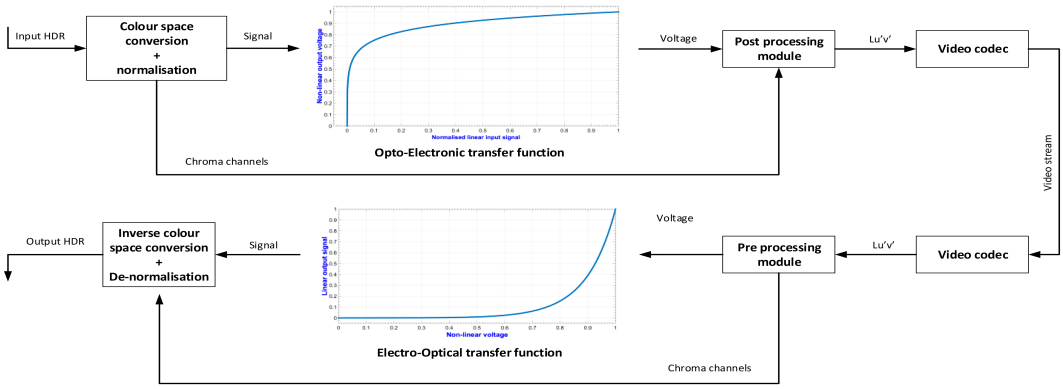


Figure 3.5: Generic schema of OETF based HDR video compression

An Opto-Electronic/Electrical transfer function (OETF) maybe defined as the relation between the input scene radiance to the output video signal value and is independent of the rendering intent [Ser11]. The primary goal of an OETF is to encode the HDR content into an 8-10 bit video signal while preserving as much scene dynamic range as possible to provide latitude for post-processing such as colour gradation etc. The non-uniform sensitivity of the HVS is used routinely by the camera manufacturers to design a non-linear TF which maps a normalised linear input signal, say $S \in [0, 1]$, to a non-linear output signal (also termed as voltage) $V \in [0, 1]$. The voltage can be disretised to an *n-bit* JND scale where n is the number of bits available. Although, OETFs take into account the just perceivable difference in brightness, unlike PTFs which typically operate on absolute scaled luminance, OETFs operate on normalised linear input signals offering the flexibility to be discretised to any available bit-depth. However, the availability of more bits results in a finer quantisation and eliminates banding or contouring artefacts.

For HDR video content, the physical luminance y is normalised such that $L = \frac{y}{N}$, where $N = max(y)$ is the normalisation factor and $L \in [0,1]$. The OETF is then applied to L resulting in a non-linearly encoded voltage signal V. This is subsequently discretised and passed to the codec for encoding along with the metadata which includes the normalisation factor. On the decoder side, an inverse function, known as Electro-Optical transfer function (EOTF) converts the non-linear voltage to normalised linear signal L. This is multiplied by the normalisation factor in order to reconstruct the output HDR signal.

For completeness, this section demonstrates the basic application of OETF and EOTF using the recommended REC 1886 standard [Ser11]:

The non-linear encoding of physical luminance to luma is shown in equation 3.17:

$$\begin{aligned} y &= 0.2126R + 0.7152G + 0.0722B \\ L &= \frac{y}{max(y)} \\ V &= \begin{cases} 4.5 \cdot L & \text{if } L < 0.018 \\ 1.099 \cdot L^{0.45} - 0.099 & \text{if } L \geq 0.018 \end{cases} \\ V_{dis} &= \lfloor V \times (2^n - 1) \rfloor \end{aligned} \tag{3.17}$$

where n is the number of allowable bit-depth. The luma to luminance conversion is shown in equation 3.18

$$\begin{aligned} V &= \frac{V_{dis}}{2^n - 1} \\ L &= \begin{cases} \frac{V}{4.5} & \text{if } L < 0.081 \\ \frac{V+0.099}{1.099} & \text{if } V \geq 0.081 \end{cases} \\ y &= L \times max(y) \end{aligned} \tag{3.18}$$

Some of the most widely used OETFs for HDR video compression are the *Perceptual Quantizer* (PQ) proposed by Dolby [PQ14, MND13] and more recently, the *hybrid log-gamma* (HLG) curve proposed by the BBC [BC15, ari15]. Also, medium-high dynamic range camera manufacturers offer proprietary version of inbuilt OETFs (mostly 10-bit). These include *Filestream* from Thomson, *S-Log* from Sony, *Panalog* from Panasonic, *Canon Log* (8-bit) from Canon and *Log C* from Arri [Stu14] which are mostly quasi-log OETFs with a knee-function [BC15]. Furthermore, a number of OETFs have been proposed as a response to Motion Pictures Experts Group (MPEG) committee's call for evidence (CfE) to standardise HDR and Wide Colour Gamut (WCG) content compression using the HEVC codec [LFH15]. Further details of the ongoing work is available in [DLCMM16b]. A generic diagram of an OETF based HDR video compression scheme is given in Figure 3.5.

3.2.3 Transfer functions in HDR video compression

PTFs and OETFs have often been for HDR video compression purposes especially by *non-backward* compatible compression algorithms. Mantiuk et al. [MKMS04] proposed an algorithm which extended the MPEG-4 encoder to support higher bit-depth encoding. The authors used the *tvi* function proposed by Ferwarda et al. [FPSG96] to derive the PTF which maps physical luminance values to an 11-bit luma space while preserving traditional 8-bit encoding of the chroma channels. Further details are given in Section 3.3.1. Similarly, Garbas and Thoma [GT11] proposed *non-backward* compatible algorithm which uses an adaptive version of the logarithmic response function [MT10] to map physical luminance to a 12-bit JND scaled luma space. More recently, Miller et al. [MND13] and Borer et al. [BC15] proposed OETFs to encode HDR video content to a 10-bit perceptually quantised JND scale to be used with 10-bit video codecs. Further details are given in Sections 3.3.4 and 3.3.5, respectively. Similar PTF/OETF based proposals have been put forward as a response to the MPEG committee's call for evidence (CfE) to explore HDR video encoding using the HEVC codec. Further details are available in [DLCMM16b]

3.3 Overview of HDR video compression algorithms

This section provides an overview of a number of the published/patented HDR video compression algorithms following either of the two approaches highlighted in Section 3.1.

3.3.1 Perception Motivated HDR video compression (*hdrv*)

Mantiuk et al. [MKMS04] proposed the first HDR video pre/post processing (compression) algorithm. According to the design specifications, the authors introduced a novel, *non-backward* compatible HDR compression scheme to extend the existing MPEG-4 video codec to encode HDR video content. The method proposes two primary modifications to the codec. First, it extends the typical 8-bit luma channel encoding in legacy video codecs to accommodate an 11-bit perceptually uniform luma code values. The authors argue that by introducing a TF, derived from the TVI function introduced by Ferwarda et al. [FPSG96], 11-bits are sufficient to encode the entire range ($y \in [10^{-5}, 10^{9}]$) of visible physical luminance. Second, it introduces a novel spatial domain horizontal edge coding of high-frequency luma components to reduce the light scattering effect. This section provides a brief overview of each module of the compression algorithm. Further details are available in [MKMS04].

Perceptually uniform quantisation of luminance

Traditional video codecs such as MPEG-4 (main profile) [AMT] are only able to encode upto 8-bits of information. Therefore they are unable to support HDR content. To overcome

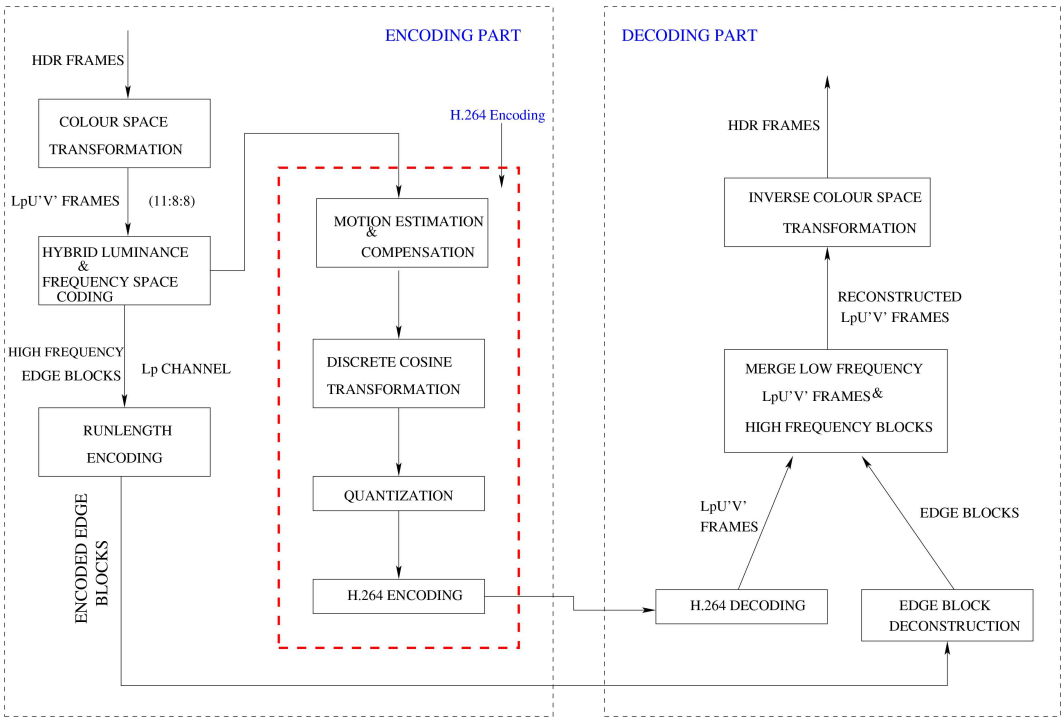


Figure 3.6: Perception-motivated HDR video encoding and decoding scheme

such a limitation, the authors introduced a novel luminance encoding scheme by virtue of which input physical luminance is encoded to an 11-bit perceptually uniform (integer) luma code values using a PTF (see Section 3.2.1). First the input luminance is calculated using the REC. 709 [Int02] primaries from calibrated linear RGB images. Subsequently, the mapping function from physical luminance to 11-bit luma space is defined as in equation 3.19.

$$\psi : L_p \rightarrow y[cd/m^2], \textit{where } L_p = [0, 2^n - 1] \tag{3.19}$$

where n in this case defines the minimum bit-depth required to encode the full range of y. The perceptually uniform luma space L_p is derived by replacing the expression $t(y_{adapt})$ in equation 3.6 by Ferwarda's TVI function [FPSG96] and forming an ordinary differential equation such that:

$$\frac{d\psi(L_w)}{dL_w} = \frac{2}{a} tvi(\psi(L_w)) \tag{3.20}$$

The boundary conditions are set such that $\psi(0) = 10^{-5}\ cd/m^2$ and $\psi(L_w, max) = 10^8\ cd/m^2$ and $a \geq 1$ represents the conservative constant. Solving the differential equation leads to a mapping function $f(\cdot)$ which maps $y \in [10^{-5}, 10^9]$ to $L \in [0, 2^{11} - 1]$. Rounding the values of L to the nearest integer results in an 11-bit perceptually encoded luma space. Further details of luminance to luma encoding is given in [MMS06, AMS08b]. The chroma encoding is retained to the 8-bit traditional representation of colours similar to LogLuv, taking into account the limitations of HVS [MKMS04].

Hybrid frequency space coding:

Subsequently, motion estimation and inter frame prediction is done as in standard MPEG-4 encoder followed by invisible noise removal in the frequency domain using Discrete Cosine Transformation (DCT). However, DCT coefficient quantisation introduces noise artefacts which are generally ignored in LDR data but pose significant issues in HDR content. To alleviate this issue the authors introduced Hybrid Frequency Space wherein sharp edges detected locally (variance in an 8x8 window) are isolated and subtracted from the original frame into high frequency edge blocks, thus keeping a resultant low-frequency frame to be encoded by the MPEG-4 encoder.

Edge block encoding:

The edge blocks segregated in the previous step, are run length encoded since most of the values are zero. Therefore, run length encoding provides a computationally efficient and straightforward technique of compressing high frequency components. The proposed hybrid block encoding improves the quality of encoded sequences albeit, at the cost of a slightly larger bitstream.

Decoding and converting to HDR frames:

The decoding is performed in three steps. Firstly, the edge map and the DCT coefficients are decoded from the bit stream. Secondly, the two channels (MPEG-4 encoded low-frequency frames & high-frequency edge blocks) are combined and finally, an inverse mapping technique from 11 bit luma to real world luminance is applied to convert perceptually quantized LuâĂŹvâĂŹ (11-8-8 bit representation) to XYZ colourspace, subsequently followed by inverse transformation to RGB (HDR frames). The schematic diagram of the encoding and the decoding block is given in Figure 3.6.

3.3.2 Non-linear encoding of HDR video content (*zhang*)

Zhang et.al [ZRB11] proposed another *non-backward* compatible HDR video encoding scheme (see Figure 3.7) which takes advantage of the higher bit-depth encoding support in existing state-of-the-art codecs (specifically, the reference H.264/AVC codec [AMT]) and follows the general methodology as shown in Figure 3.1a producing a single optimally quantised bit stream of encoded HDR video data.

Colour space transformation and optimal bit-depth quantisation:

The first step involves converting floating point linear RGB frames to the 32-bit LogLuv format (see Section 2.3 with 16-bit integer representation of luminance and 8-bit integer representation of the chroma channels. However, as explained previously, existing codecs

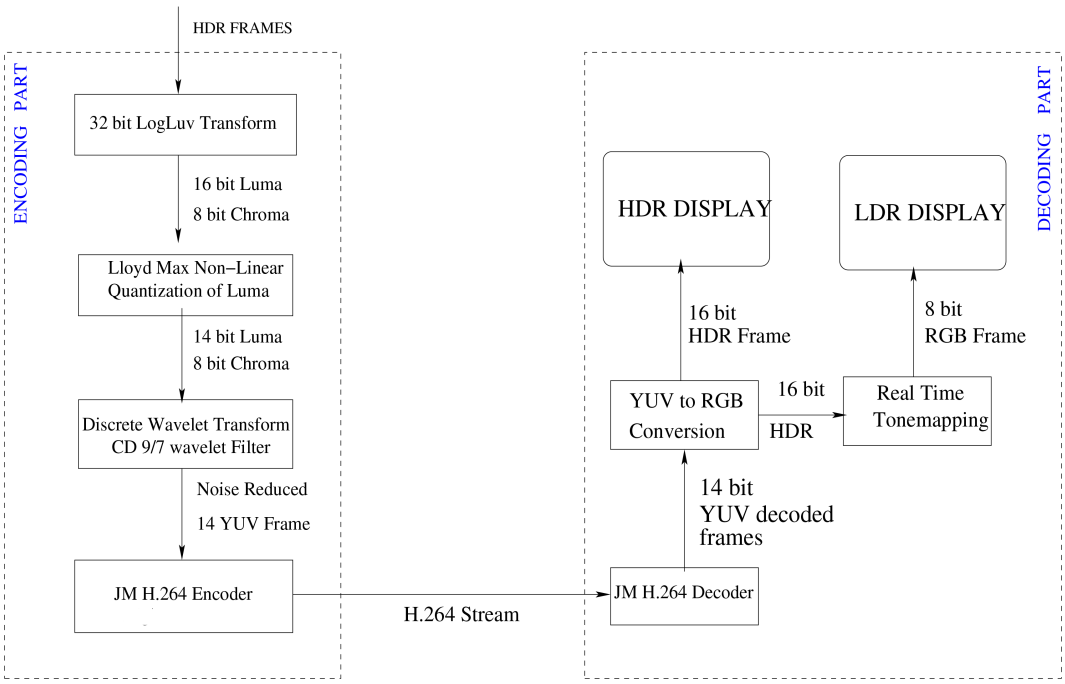


Figure 3.7: Encoding and Decoding scheme of HVS based optimal bit-depth HDR video compression

can support up to 14 bits/channel integer representation. Therefore, to avoid clamping and losing higher luminance information, the authors introduced a novel non-linear quantisation scheme analogous to the Lloyd Max quantisation technique [Sch96] in order to quantise the 16-bit luminance channel y to a 14-bit luma channel L_p. The HVS based non-linear quantisation uses a contrast sensitivity function [ZRB11] to iteratively weigh the information to be preserved in the 14-luma space such that the loss in luminance information is $< 1JND$. The chroma channels u', v' are discretised to 8-bits similar to LogLuv. The $14:8:8$ $L_p u' v'$ frame is then passed to the noise reduction module for invisible noise correction.

Noise reduction of 14-bit frames:

The quantisation of the luminance channel from 16-bit to 14-bit integer presentation leads to noise artefacts which appear in the 14-bit non-linearly quantised luma channel. The resultant noise is filtered using state-of-the-art CD 9/7 wavelet filter pair [DS98, Swe98]filtered before the 14-bit frames are passed to the codec.

Encoding 14 bit frames:

State-of-the-art codecs such as the reference H.264/AVC encoder [AMT, WSBL03] support up to 14 bits/channel encoding. Therefore, after invisible noise filtering the 14 bit *.yuv* files are generated and passed on to the encoder. The frames are encoded in High 4:4:4 profile with an I-P-P-P GOP (group of pictures) structure.

Decoding and conversion to HDR frames:

The 14-bit HDR stream is decoded and 14-bit YUV frames are converted to HDR frames using inverse LogLuv which and subsequently displayed using an HDR display [SHS*04]. In parallel, the HDR frames are tone mapped using any real-time global TMO and displayed as LDR stream as shown in Figure 3.7.

3.3.3 Temporally Coherent Luminance to Luma mapping (*fraunhofer*)

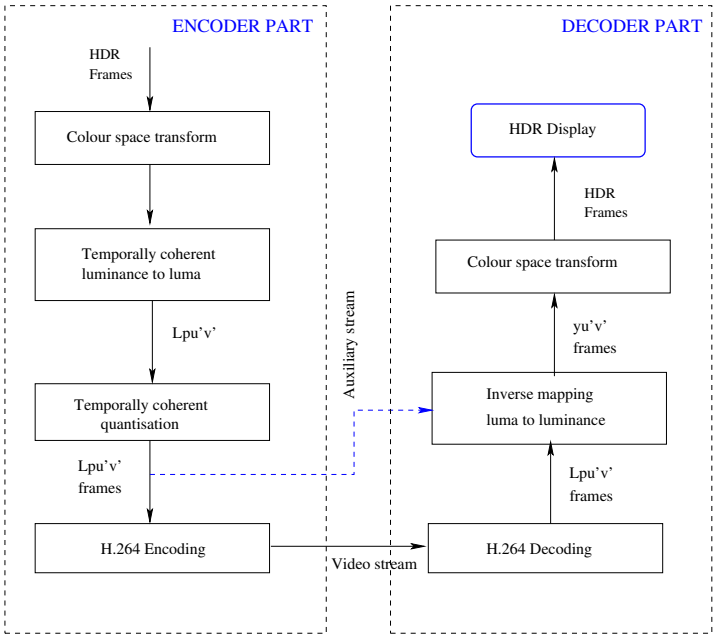


Figure 3.8: Schematic diagram of Temporally Coherent Luminance to Luma mapping

Garbas & Thoma [GT11] proposed a new HDR video encoding technique in 2011. This too is a non-backward compatible HDR video encoding technique which takes advantage of the higher bit-depth support in existing state-of-the-art H.264/AVC encoders. Figure 3.8 shows the general encoding and decoding scheme. As mentioned previously in Section 2.3.4, the authors proposed a modification of the LogLuv encoding. The LogLuv encoding technique maps real world luminance in the interval of $[5.44 * 10^{-20}, 1.84 * 10^{19}]$ to 15-bit integer luma values in the interval of $[0, 2^{15} - 1]$. However, the dynamic range covered by LogLUV mapping is far beyond the range of what the HVS can simultaneously perceive. Therefore, the authors argue that reserving bits to represent imperceivable luminance values is redundant and would degrade compression efficiency. A scaling factor was introduced in [MT10] to scale individual frames in video sequence to exploit the entire range luma values for a given bit depth.

However, this was a relatively straightforward extension to LogLuv encoding which lacked specific video encoding aspects, most notably introducing the possibility of severe

flickering artefacts due to scaling without taking temporal coherence of successive frames (in a video sequence) into account. In [GT11], the temporal coherence of successive frames is taken into account to extend the adaptive mapping of captured HDR frames to a 12-bit luma and two 8-bit chroma channels.

Temporally coherent Luminance to 12 bit Luma mapping

In this stage of the pipeline the RGB (HDR) frames are first converted to Yu'v' frames with 8-bits/pixel allocated to each of the chroma channels u' & v'. The Y channel contains the real-world luminance captured in a particular frame. Following Yu'v' conversion, the luminance values are subsequently mapped to 12-bit luma values. As previously mentioned in Section 2.3.4, the non-adaptive luminance-to-luma mapping can be defined as

$$L_n = \lfloor \frac{2^n - 1}{\log_2(Y_{max}/Y_{min})} (\log_2(Y) - \log_2(Y_{min})) \rfloor, \tag{3.21}$$

$$Y = 2^{(L_n + 0.5)\frac{\log_2(Y_{max}/Y_{min})}{2^n - 1} + \log_2(Y_{min})} \tag{3.22}$$

where L_n is the mapped luma value from real-world luminance, Y is the luminance value of each pixel, $[Y_{min}, Y_{max}]$ are the minimum and maximum frame luminance respectively and n is the representable luma bit-depth.

However, the dynamic range of a scene can vary between successive frames which would introduce severe flickering due to non-adaptive mapping and prevent temporal prediction during H.264 encoding. Therefore, the non-adaptive mapping is extended to take temporal coherence into account and can be defined as

$$L_{n,l} = (L_{n,k} + 0.5)\frac{\log_2(Y_{max,k}/Y_{min,k})}{\log_2(Y_{max,l}/Y_{min,l})} + (2^n - 1)\frac{\log_2(Y_{min,k}/Y_{min,l})}{\log_2(Y_{max,l}/Y_{min,l})} \tag{3.23}$$

where $L_{n,l}$ and $L_{n,k}$ are the mapped luma values of two successive frames k and $l = k + 1$. Therefore, taking temporal coherence into account while mapping luminance-to-luma values allows for perfect temporal prediction where the dynamic range of the scene changes abruptly, thus reducing the effects of flickering. Further

Temporally coherent quantisation

Due to temporal prediction, different luminance ranges are mapped to different luma values. Therefore, fixed quantisation according to identical quantization parameter in H.264/AVC leads to varying quantisation of the luma channel depending on the mapping. Therefore, the authors propose to take luminance mapping range into account for each frame in order to find a suitable quantization parameter (QP) value accordingly. The QP value of the 1^{st} frame is taken as reference QP and subsequent changes in quantization values ΔQP are

calculated accordingly. The relative quantization according to mapping range is defined as

$$Q_{rel_{l,k}} = \frac{Q_{step,l}}{Q_{step,k}} = \frac{\log_2(Y_{max,k}/Y_{min,k})}{\log_2(Y_{max,l}/Y_{min,l})} \tag{3.24}$$

According to the definition Q_{step} approximately doubles when QP values is increased by 6 units ΔQP is calculated as

$$\Delta QP = round(6.\log_2(Q_{rel_{l,k}})) \tag{3.25}$$

and any arbitrary frame l will be quantized with QP value

$$QP_l = QP_1 + \Delta QP_{l,1} \tag{3.26}$$

3.3.4 Perceptually quantised HDR video compression (*PQ*)

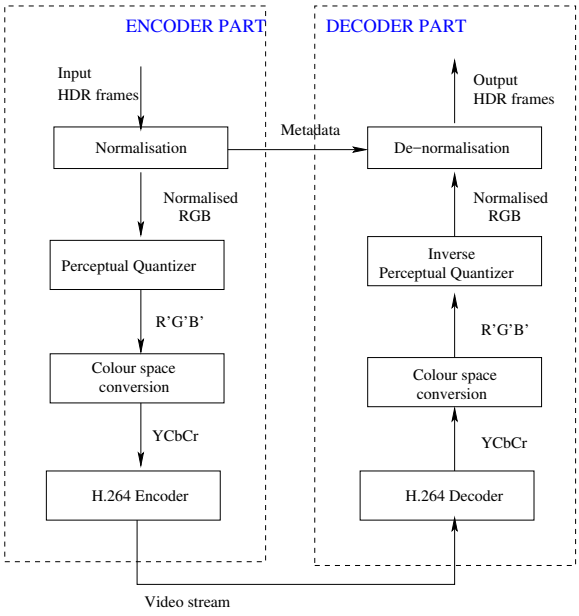


Figure 3.9: Schematic diagram of the Perceptual Quantizer compression algorithm.

Based on the HVS contrast sensitivity model developed by Barten et al. [Bar92], Miller et al. [MND13] proposed an OETF-EOTF for HDR video compression purposes. The proposed OETF-EOTF has recently been standardised as the *Society of Motion Picture & Television Engineers* (SMPTE) standard 2084 [PQ14] for encoding HDR and Wide Colour Gamut (WCG) content. Although, the proposed OETF is based on Barten's CSF model, unlike the GDF (see Section 3.2.1) used for medical imaging purposes, this OETF maps a larger range of pixel values $V \in [10^{-4}, 10^4]$ cd/m^2 to a 10-bit JND inspired code value range of $L \in [0, 2^{10} - 1]$. According to the authors, the proposed non-linear OETF-EOTF ensures the optimal usage of available code values (bit-depth) to map physical lumi-

nance to a 10-bit luma range. The HDR video compression algorithm designed on the basis of this OETF and EOTF is described as follows:

Compression

Input HDR frames are first linearly normalised such that the pixel values $V \in [0,1]$. The proposed non-linear OETF as mentioned in equation 3.27 is then applied to V to non-linearly encode the individual channels such that the resultant non-linear signal $\mathrm{L} \in [0,1]$.

$$L = \left(\frac{c_2 V^{m_1} + c_1}{c_3 (1 + V^{m_1})} \right)^{m_2} \tag{3.27}$$

where V represents the normalised pixel values and L represents the normalised non-linearly encoded code values. The constant values are given in Table 3.1. Subsequently, the non-

$m_1 = 0.15930$	$m_2 = 78.84375$	$c_1 = 0.83593$	$c_2 = 18.85156$	$c_3 = 18.6875$

Table 3.1: Table of constants used by the Perceptual Quantizer based signal encoding.

linear signal undergoes colour space transformation and is discretised to 10-bits such that before being passed on to the video codec to produce a 10-bit output video stream. The normalisation factor is typically the maximum pixel value of each input HDR frame and is stored as auxiliary metadata to be used later for decompression purposes later.

Decompression

On the decompression side, the bitstream is decoded and individual frames undergo normalisation and an inverse colour space transform. The EOTF function mentioned in equation 3.28 is then applied to L and the resultant is subsequently multiplied by the normalisation factor of each frame obtained from the auxiliary metadata in order to reconstruct the decoded HDR frames.

$$V = \left(\frac{L^{\frac{1}{m_2}} - c_1}{c_2 - c_3 L^{\frac{1}{m_2}}} \right)^{\frac{1}{m_1}} \tag{3.28}$$

The constants in this case are mentioned earlier in Table 3.1.

3.3.5 Hybrid log-gamma based HDR video compression (*hlg*)

Borer et al. [BC15, ari15] proposed an OETF-EOTF (see Section 3.2.2) based HDR video compression algorithm derived from the REC. 709 [Int02] gamma function. Similar to the PQ (see Section 3.3.4), this OETF is influenced by the HVS perception to brightness. However, unlike the PQ, the primary working principle of this compression algorithm is that the same content can be viewed on LDR and HDR displays without further adaptation. As explained previously in Section 3.2, the HVS perception to brightness can be modelled

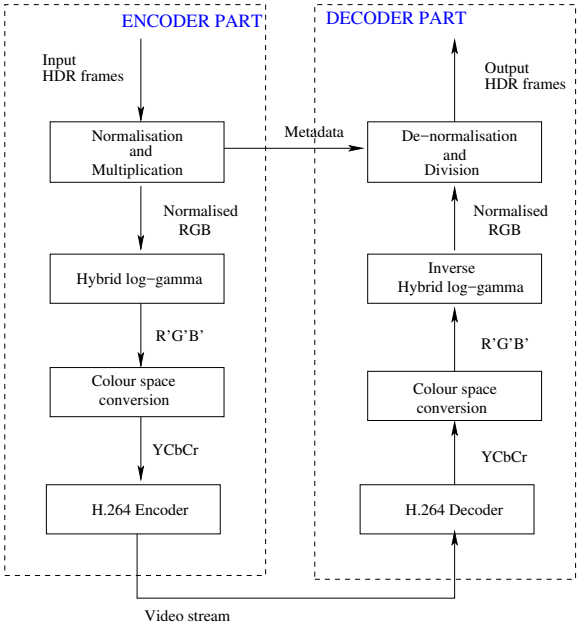


Figure 3.10: Schematic diagram of the hybrid Log-Gamma compression algorithm.

using the *Weber's Law* for brighter regions. This means that a logarithmic TF can be used to model the HVS perception. For darker regions however, the HVS perception can be more accurately modelled using the *De-Vries Rose* relationship which is similar to the gamma non-linearity used in REC. 709 to model dimmer (more specifically Cathode Ray Tube (CRT) i.e. peak brightness ≤ 100 cd/m^2) displays. This means that for darker regions, the HVS perception can be modelled by a *1/2 power* model. Therefore, in order to display a larger range of luminance from dark to very bright regions, an OETF can be designed by merging the *1/2 power* model and the logarithmic TF. The *hybrid log-gamma* OETF proposed by Borer et al. [BC15] follows a similar approach as shown in equations 3.29.

$$L = \begin{cases} r\sqrt{V} & \text{if } V \in [0,1] \\ a \cdot log(V-b)+c & \text{if } V > 1 \end{cases} \tag{3.29}$$

where L is the non-linear output response signal, V is the linear input signal, r is the reference output signal value and a, b and c are defined such that the $L = 1$ when $V = 12$. The values of r, a, b and c are given in Table 3.2.

$r = 0.5$	$a = 0.17883277$	$b = 0.28466892$	$c = 0.55991073$

Table 3.2: Table of constants used by the hybrid log-gamma OETF

Similarly, the EOTF can be derived as shown in equation 3.30.

$$V = \begin{cases} (\frac{L}{r})^2 & \text{if } L \in [0, r] \\ e^{(\frac{L-c}{a})} + b & \text{if } L > r \end{cases} \tag{3.30}$$

The HDR video compression algorithm designed on the basis of the above mentioned OETF and EOTF is described as follows:

Compression

Input HDR frames are first linearly normalised such that the pixel values $V \in [0,1]$. The normalised pixel values are then linearly multiplied by an arbitrary constant of 12.0 such that $V \in (0,12]$. The OETF mentioned in equation 3.29 is then applied to V such that the output non-linear signal $L \in [0,1]$. Subsequently, the pixel values undergoe colour space conversion and discretisation before being passed on to the codec and encoded at 10 bits/pixel/channel. The normalisation factor, typically the maximum value of each HDR frame is stored as look-up table and passed on as an auxiliary metadata stream which is used by the decompression side of the algorithm.

Decompression

The output stream undergoes a reverse process whereby the decoded YC_bC_r frames are converted to RGB' and subsequently the EOTF as given in equation 3.30 is applied to the signal. The resultant RGB is then normalised by a factor of 12.0 and subsequently multiplied by the normalisation factor obtained from the look-up table, thus reproducing the decoded output HDR frames.

3.3.6 Backward compatible HDR-MPEG (*hdrmpeg*)

Mantiuk et al. [MEMS06] proposed the first *backward compatible* HDR compression algorithm. This method incorporated backward compatibility as shown in Figure 3.1b by creating a tone mapped base stream which can be played back on an LDR screen using any available video player. However, the method also introduces a new colour space transformation, a reconstruction function which can be considered as a precursor to inverse tone-mapping and non-linear quantisation. The steps are described as follows:

The LDR stream:

Considering backward compatibility with existing 8 bit video decoders, the HDR video content is tone mapped, using photographic TMO, to produce an 8 bit RGB frames. They are transformed to YCbCr colour space and encoded using any MPEG-4 encoder. The LDR stream can be played back on any LDR displays in the absence of an HDR display.

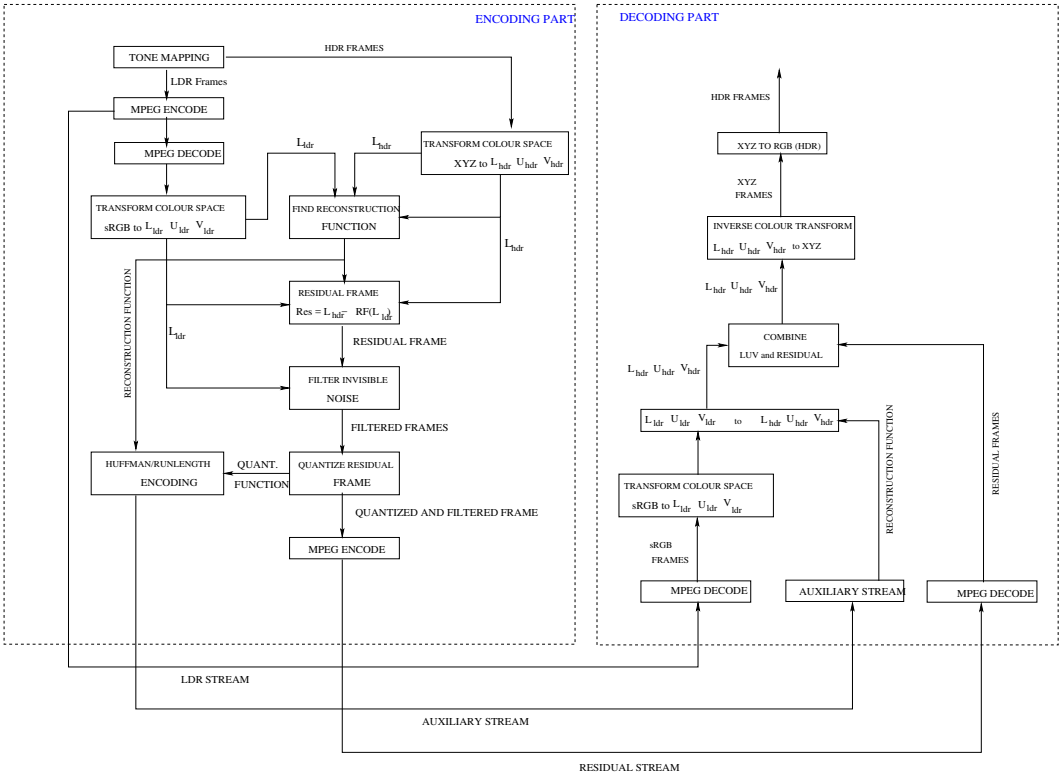


Figure 3.11: Schematic diagram of HDR-MPEG

Colour space transformation:

The method introduces a new backward compatible perceptually uniform Lu'v' colourspace [Man06] which is able to encode Luminance values of HDR as well as LDR frames. This is done to ensure that colour channels of both LDR as well as HDR pixels contain the same information. The decoded LDR frames are transformed from gamma corrected sRGB to $L_{ldr}U_{ldr}V_{ldr}$ and the corresponding HDR frames are transformed to $L_{hdr}U_{hdr}V_{hdr}$ colour space with 12 bits allocated to encode the luminance and 8 bits each for two chroma channels. To encode real world luminance Y to 12 bit luma, l_{hdr}, the following conversion formula is used:

$$l_{hdr}(y) = \begin{cases} a.y & \text{if } y < y_l \\ b.y^c + d & \text{if } y_l \leq y < y_h \\ e.\log(y) + f & \text{if } y \geq y_h \end{cases}$$

and the inverse operation to map l_{hdr} to real world luminance y are:

$$y(l_{hdr}) = \begin{cases} a'.l_{hdr} & \text{if } l_{hdr} < l_l \\ b'(l_{hdr} + d)^c & \text{if } l_l \leq l_{hdr} < l_h \\ e'.\exp(f'.l_{hdr}) & \text{if } l_{hdr} \geq l_h \end{cases}$$

The constants are given in the table below: Henceforth, all operations are conducted

a =17.554	e = 209.16	a' = 0.056968	e' = 32.994
b = 826.81	f = -731.28	b' = 7.3014e-30	f' = 0.0047811
c = 0.10013	yl = 5.6046	c' = 9.9872	ll = 98.381
d = -884.71	yh = 10469	d' = 884.17	lh = 1204.7

Table 3.3: Constants used for the Luminance and Luma mapping

on the luma channel.

Reconstruction function:

The authors introduce a strictly monotically increasing reconstruction function using a look-up table (LUT). This used to predict HDR pixel values from its corresponding LDR frame. The reconstruction function essentially maps LDR pixel values to HDR pixel values, contained in one of the 256 bins of the LDR pixel values. It is defined as the arithmetic mean of the all the pixels in a particular bin Ω_i and is given in equation 3.31.

$$RF(l) = \frac{1}{Card(\Omega_i)} \Sigma l_{hdr}(i) \;\; where \;\; \Omega_i = i \in [1,N] : l_{ldr} = l \tag{3.31}$$

$l \in [0,255]$ is an index of a bin, N is the spatial resolution of a frame, $l_{ldr}(i)$ and $l_{hdr}(i)$ are luma values of the i-th LDR and HDR pixel respectively.

Residual frame computation:

The reconstruction function (lookup table) as mentioned in Section 3.3.6 is then used to predict the L_{hdr} values from the $L_{ldr}values$ resulting in a $Predicted_{hdr}$ luma frame. Subsequently, the residual luma is calculated as:

$$Residual_l = L_{hdr} - prLuma_{hdr} \text{ where } prLuma_{hdr} = RF(L_{ldr}) \tag{3.32}$$

Therefore, the accuracy of the predicted L_{hdr} and $Residual_l$ largely depends on the accuracy of the reconstruction function.

Noise Reduction and frame quantisation:

Residual rames do not compress well primarily because they contain a lot of high frequencies including noise. To mitigate this problem, invisible noise filtering is applied to the residual frame using the CDF 9/7 discrete wavelet filter pair [XWHL94, WL05].

The filtered frame can ideally contain values up to 12 bits (0 to 4095) which cannot be encoded using an 8-bit encoder. The authors introduce a simple yet effective quantization

function to quantize and limit residual pixel values to 8 bits as given below,

$$\hat{Res}_l(i) = [Res_l(i)/q(m)]^{-}127 \div 127, \text{where } m = k \Leftrightarrow i \subset \Omega_k \tag{3.33}$$

and the quantization factor, q(m), is calculated for each bin Ω_k as:

$$q(m) = max(q_{min}, \frac{max_{i \in \Omega_i}(|Res_l(i)|)}{127}) \tag{3.34}$$

The quantization factors $q(m)$where $m \in [0,255]$ is stored in the auxiliary stream alongside the reconstruction function. The entire encoding is visually described in Figure 3.11.

Decoding and merging to HDR:

The decoding process is fairly straightforward. The decoded sRGB frames are transformed to $L_{hdr}u_{hdr}v_{hdr}$ colourspace. Using the reconstruction function from auxiliary stream, $L_{hdr}u_{hdr}v_{hdr}$ values are predicted and finally merged with the decoded residual frames Res_l to re-create the HDR frame. The hybrid luma space is inverse mapped to real world luminance (see Section 3.3.6) and Yu'v' is transformed to XYZ followed by inverse transformation to 16 bit RGB frames.

3.3.7 JPEG-HDR for video (*hdrjpeg*)

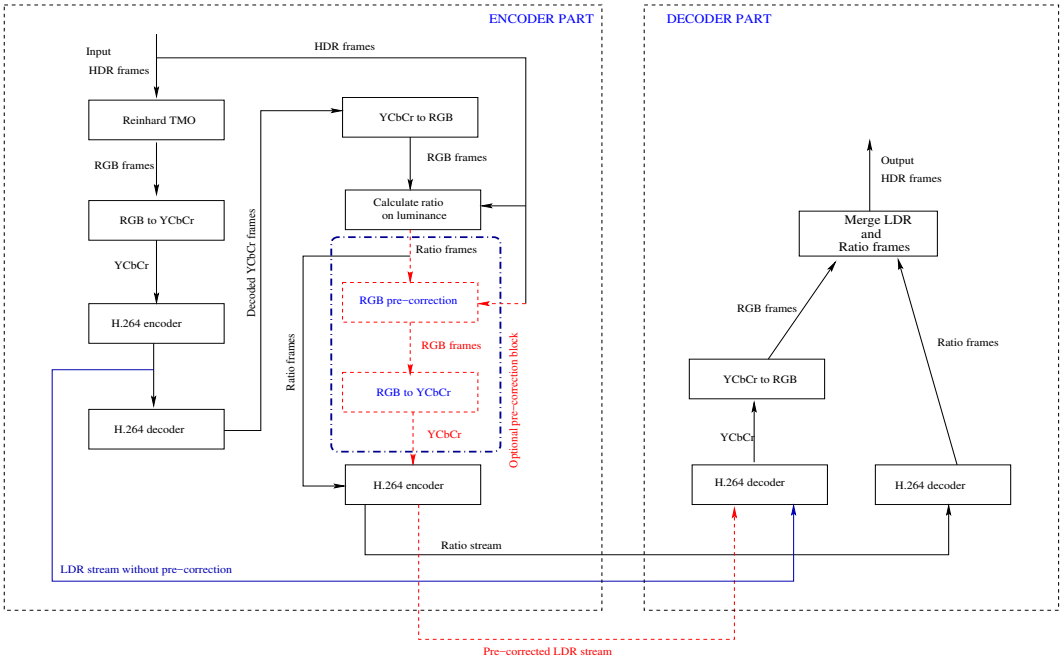


Figure 3.12: Schematic diagram of JPEG-HDR based video encoding (with optional post-correction).

As previously mentioned in Section 2.2.3, Ward and Simmons [WS06] proposed

a static HDR image compression algorithm which uses the Photographic TMO to create the primary LDR frame and the residual is calculated by the ratio between the reference HDR frame luminance and the luminance of the tone-mapped LDR frame. This is then subsampled and stored as subband of the LDR frame. On the decoder side, the ratio frame is upsampled and merged with the LDR frame to create the decoded HDR. Further details about the procedure is given in [WS06]. However, it is interesting to note that it can be fairly straightforward to adapt this HDR image compression algorithm for video compression purposes. The video implementation of this was carried out for the work presented later in Chapter 6.

Base stream

Similar to *hdrmpeg* [MEMS06], the *hdrjpeg* (video) algorithm is a *backward* compatible compression algorithm which follows a two-pass encoding scheme. On the first pass, input HDR frames are tone-mapped using the photographic TMO to create the primary LDR stream. A major change for video adaptation is the replacement of the photographic TMO [RSSF02] with a temporally coherent version of the photographic TMO [KRTT12]. The temporally coherent version reduces the flickering artefacts which might have been introduced by using a static image TMO in a video sequence. The LDR frames are then gamma corrected and changed to an codec suitable colour space (RGB to YC_bC_r) before being passed on the codec to create the *base* stream.

Detail stream

On the second pass, the LDR stream is decoded and the luminance of the input HDR frames as well as the decoded LDR frames are calculated using the REC. 709 primaries [Int02]. Subsequently, a ratio frames are created by dividing the HDR luminance by the decoded LDR luminance and log-encoded as shown in equation 3.35

$$ratio_{(x,y)} = \log(\frac{h_{(x,y)}}{l_{(x,y)} + \varepsilon}) \tag{3.35}$$

where $\text{ratio}_{(x,y)}, \text{h}_{(x,y)}$ and $\text{l}_{(x,y)}$ are the pixel values of the ratio frame, the HDR luminance and the LDR luminance, respectively. A negligible constant ε is added to the denominator to avoid divide-by-zero conditions. Unlike the original algorithm, the video implementation does not sub-sample the ratio frame and store the same as a sub-band of the LDR image. In order to preserve maximal video quality, the ratio frames are encoded as a separate *detail* stream in full resolution with a 4:0:0 sub-sampling format (see Section 3.4.4).

Similar to the original algorithm which employs pre- or post-correction techniques to retain image quality, an optional pre-correction is implemented in the video implementation. In the pre-correction block, the ratio frames derived earlier is used to correct the

base (RGB) stream and reduce the distortions introduced to the base stream during the first encoding pass as shown in equation 3.36.

$$RGB_{(x,y,z)} = \frac{HDR_{(x,y,z)}}{Ratio_{(x,y)}} \tag{3.36}$$

where $\mathrm{RGB}_{(x,y,z)}$, $\mathrm{HDR}_{(x,y,z)}$ and $\mathrm{Ratio}_{(x,y)}$ are the pre-corrected LDR-RGB, reference HDR-RGB and derived Ratio frames, respectively. Subsequently, the pre-corrected LDR frames and Ratio frames are passed on to the codec for encoding. It is however important to note that in case the optional pre-correction block is not applied, the LDR frames encoded in the first pass is considered as the base stream.

Decoding and merging to HDR

On the decoder side, the LDR frames from the base stream with/without pre-correction undergoes colour space conversion ($\mathrm{YC_bC_r}$ to RGB) and inverse gamma-correction. The linear RGB frames are then merged with the decoded ratio frames to obtain the output HDR frames. A visual description of the algorithm in given in Figure 3.12.

3.3.8 Rate-Distortion optimised HDR video compression (*rate*)

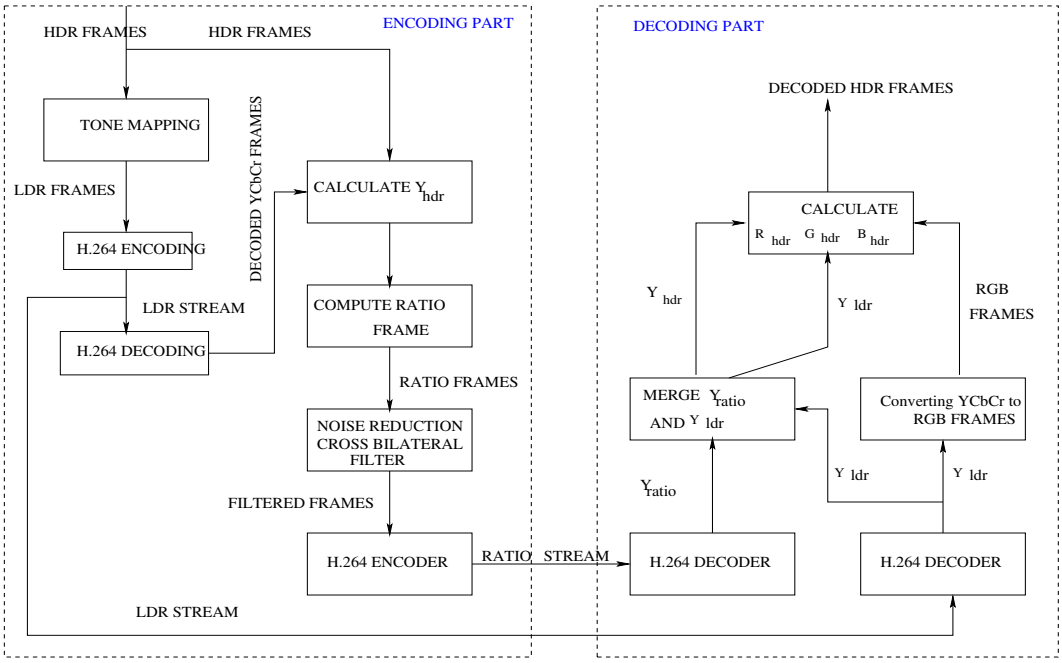


Figure 3.13: Schematic diagram of Rate-Distortion optimised HDR video encoding.

Lee et.al [LK08] proposed a backward compatible HDR video encoding scheme in 2008. Similar to the generic approach as shown in Figure 3.1b, this method also proposes two streams, the primary stream being a tone mapped LDR stream and the secondary

residual stream. The general methodology applied in this scheme is as follows:

The LDR stream:

HDR frames are tone mapped using a novel temporally coherent TMO [LK07] based on the Gradient Domain TMO [FLW02]. The temporally coherent TMO reduces severe flickering artefacts by taking advantage of motion information. The tone mapped LDR stream is encoded into an 8 bit stream using the JM.H.264/AVC encoder.

The Ratio stream:

The ratio stream represents the residual data between the HDR and LDR pixel values. Its derived by taking the ratio between uncompressed HDR and its corresponding tone mapped LDR frame on a logarithmic scale as given in equation 3.37:

$$ratio(x,y) = \log(\frac{h(x,y)}{l(x,y)+\varepsilon}) \tag{3.37}$$

Noise reduction and encoding of ratio frames:

The calculation of ratio frames leads to noise artefacts due to quantization. A cross-bilateral filter [TM98] is applied to individual frames to remove quantisation noise while preserving sharp edges. Subsequently, the ratio frames are encoded using the JM.H.264 encoder in 8-bit high profile mode.

However, the bit-rate of the LDR and Ratio streams are not similar. The quantization parameters of the LDR (QP_{LDR}) and Ratio (QP_{ratio}) streams are controlled such that distortions of the reconstructed HDR steam is minimized. The optimization can be solved by minimizing the Lagrangian cost function, given by:

$$J = D_{LDR} + \mu D_{HDR} + \lambda(R_{LDR} + R_{ratio}) \tag{3.38}$$

where R_{LDR} and R_{ratio} are the bit-rates of LDR and ratio streams respectively.

Decoding and merging to HDR:

Figure 3.13 describes the decoding and merging process of YC_bC_r (LDR frames) and Y_{ratio} (ratio frames) into Y_{hdr}, followed by the calculation of $R_{hdr}, G_{hdr}, B_{hdr}$ as given by,

$$\begin{bmatrix} h_r \\ h_g \\ h_b \end{bmatrix} = \begin{bmatrix} h_y(\frac{l_r}{l_y})^{1/s} \\ h_y(\frac{l_g}{l_y})^{1/s} \\ h_y(\frac{l_b}{l_y})^{1/s} \end{bmatrix}$$

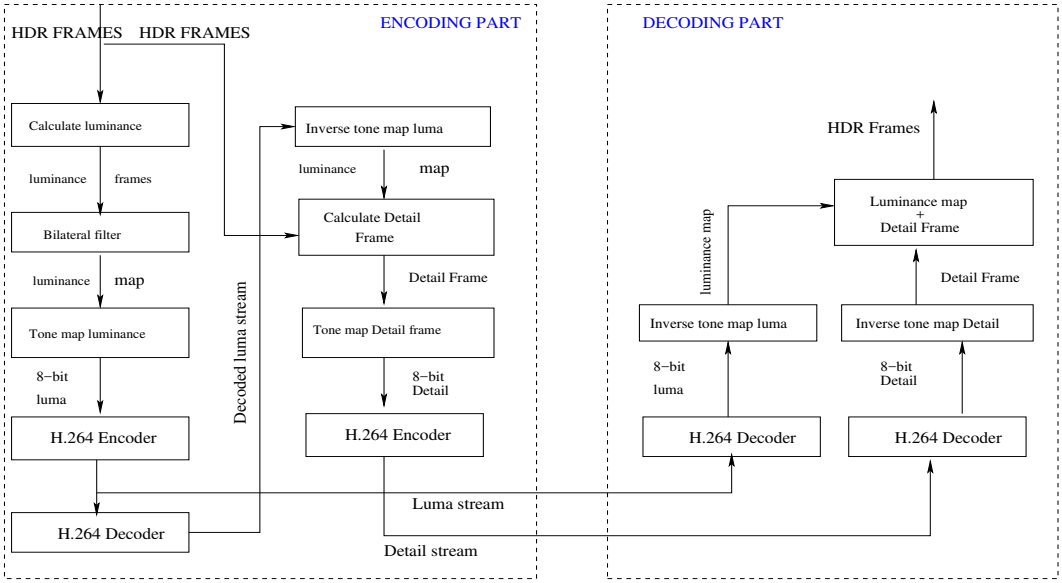


Figure 3.14: Schematic diagram of the goHDR compression algorithm.

3.3.9 HDR video data compression (*goHDR*)

Another *backward-compatible* commercial solution to HDR video compression was proposed by goHDR [CEB*10]. The compression algorithm produces a Base and a Detail streams which are both tone mapped using a sigmoidal tone reproduction operator to produce two 8-bit streams. The generic philosophy of this algorithm is to segregate the two streams based This section provides a brief overview of the algorithm.

Base stream

To create the base stream of the algorithm, the luminance y of the input HDR frame is calculated using the REC. 709 primaries [Int02]. However, to increase the compressibility of the base stream, the HDR lumimance is passed through an edge-preserving bilateral filter [TM98] and the filtered luminance y_{fil} is tone mapped using the photographic TMO (as described earlier in Section 2.5.5) where the 8-bit luma L is calculated as:

$$L = \frac{y_{fil}}{1 + y_{fil}} \tag{3.39}$$

Subsequently, the 8-bit tone mapped luma frames are passed to the video codec to create the base video stream.

Detail stream

Similar to other *backward* compatible HDR video compression algorithms, the goHDR algorithm also follows a dual-loop encoding scheme whereby the encoded base stream is

passed to the video decoder and the decoded base stream is used to create the *detail* stream. In this case, the decoded 8-bit luma frames are inverse tone mapped using an inverse sigmoidal tone mapping operator to obtain the filtered luminance as shown in equation 3.40

$$y_{fil-dec} = \frac{L}{1-L} \tag{3.40}$$

The input HDR frames are then divided by the decoded and inverse tone mapped luminance to create a 3-channel detail stream which contains the high frequency details such that:

$$Det_{frame} = \frac{HDR}{y_{fil-dec}} \tag{3.41}$$

Subsequently, similar to the base frame, the Det_{frame} is tone mapped using the sigmoidal TMO and the tone mapped frames are passed to the video codec to be encoded as the detail stream.

Decoding and merging to HDR

On the decoder side, the base and the detail encoded video stream are decoded and both streams are inverse tone mapped to obtain the $y_{\text{fil}-\text{dec}}$ and $\text{Det}_{\text{frame}}$, respectively. Subsequently, the frames are multiplied in order to obtain the decoded HDR frame as shown in equation 3.42.

$$HDR_{dec} = y_{fil-dec} \times Det_{frame} \tag{3.42}$$

3.3.10 Optimal exposure based HDR video compression (*optimal*)

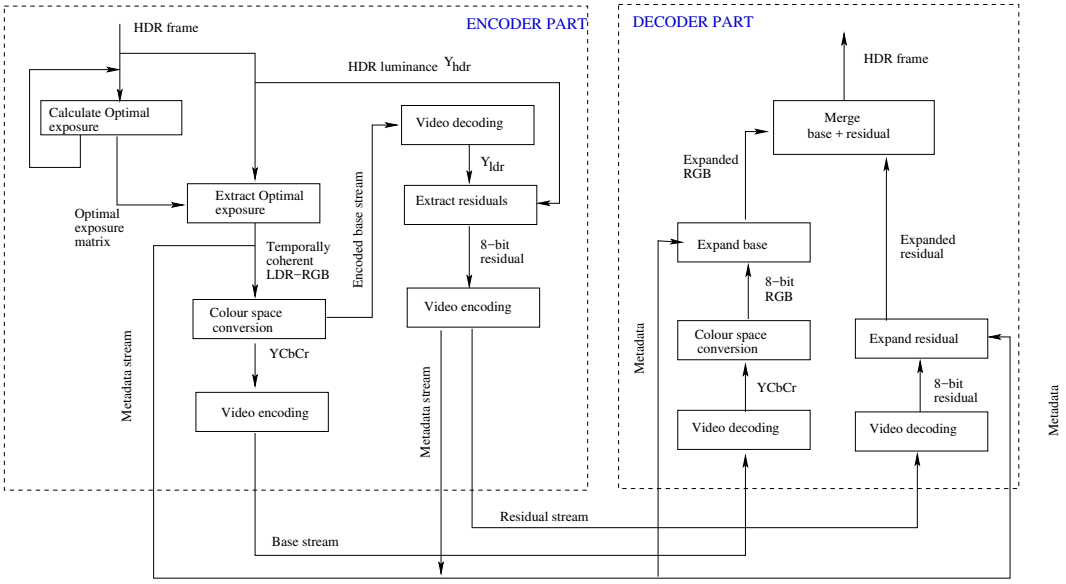


Figure 3.15: Schematic diagram of optimal exposure based HDR video compression.

Following a compression scheme similar to *hdrjpeg, hdrmpeg* and *gohdr*, Debattista et al. [DBRS*15] proposed a *backward* compatible HDR video compression algorithm which splits the input HDR frames into two LDR streams, namely, base and residual, each 8-bit of information. The 8-bit RGB base stream comprises of the optimal exposure [HW10] which can be extracted from the HDR frame. Following a dual-loop encoding technique, the base stream is encoded and decoded back to create the residual stream. This process accounts for the distortions introduced to the base frames by the codec at specific quality levels allowing the residual frames to compensate for the distortions. Subsequently, the residual frames are also tone compressed to 8-bits and encoded by the video codec. On the decoder side, both the base and the residual streams are decoded, expanded and merged to form the output HDR frames. This section provides an overview of each step of the compression algorithm and a visual description is given in Figure 3.15.

Optimal exposure extraction

The optimal exposure can be defined as the largest contiguous region of logarithmically encoded luminance which can be fitted within the allowable bit-depth of 8-bits/pixel/channel, typically supported by legacy codecs such as MPEG-2 [IIJ94]. This ensures that the maximum possible scene information is stored in a single optimally chosen exposure akin to the zone-metering system in analogue film photography introduced by Ansel Adams [Ada81, Ada83]. The logarithmic domain is chosen to match the HVS response to physical luminance. The authors argue that this technique introduces a new method to map HDR to LDR and is an alternative to tone mapping. The largest contiguous area is mathematically defined as:

$$\underset{\mathbf{E}}{maximize}(f(I(E))) \tag{3.43}$$

where $f(\cdot)$ counts the number of well exposed pixels in an HDR image I at a specific exposure E [DBRS*15]. To perform the optimal exposure extraction, the luminance of the input HDR frames y_{hdr} are first calculated using the REC 709 [Int02] primaries along with the DR of the input frame. The number of bins to create the histogram of the 8-bit LDR luminance y_{ldr} is calculated by the Freedman Diaconis rule [FD81]. The function $IQR(\cdot)$ calculates the inter-quartile range and starting from the first bin the value of all the bins within a given range is checked. Similar to a *greedy-process*, the current maximum value is stored as *best*. The process recursively iterates through all the bins and values greater than current maximum replaces the maximum value. Finally, this recursive process yields the entire histogram. The starting point of the histogram i.e. the minimum luminance value l_{min} is stored as metadata as it is sufficient to identify the range of luminance that the optimal exposure range. Optionally, the maximum luminance can be either stored or calculated by taking into account the l_{min} and the bit-depth. Further details about the optimal exposure calculation is given in [HW10] and [DBRS*15].

Base stream

This work uses a dual-loop technique to create the base stream. On the first loop, the optimal exposure of each input HDR frame is calculated from y_{hdr} and l_{min} and DR of the frame is stored as metadata. On the second loop, the obtained metadata information is used to extract the optimal exposure (luminance only) y_{opt} of each input HDR frame such that $y_{opt} \approx y_{ldr}$. Subsequently, the chroma information is extracted such that:

$$chroma = \frac{HDR_{ref}}{y_{hdr}} \tag{3.44}$$

The chroma is then multiplied by y_{opt}, gamma corrected and discretised to 8-bits to obtain the LDR base frame such that:

$$base_{rgb} = (chroma \times y_{opt})^{\frac{1}{\gamma}} \times (2^8 - 1) \tag{3.45}$$

Subsequently, the $base_{rgb}$ frames undergo colour space transformation and are converted to 8-bit YC_bC_r before being passed on to the codec for encoding. It is to be noted that authors state that the recursive calculation of the optimal exposure per frame before extraction allows for temporal coherence in the base stream.

Residual stream

Unlike the previously described *backward* compatible algorithms, this compression algorithm does not used the dual-loop encoding scheme. Instead the residual stream is calculated as:

$$y_{res} = y_{hdr} - y_{opt} \tag{3.46}$$

The y_{res} are then shifted by an $offset = 1$ and then logarithmically encoded as:

$$l_{res} = log_{10}(y_{res} + 1 - min(y_{res})) \tag{3.47}$$

The l_{res} is then normalised by a logarithmically computed factor such that:

$$\begin{aligned} l_{max} &= log_{10}(max(y_{res}) + 1 - min(y_{res})) \\ l_{res_{norm}} &= \frac{l_{res}}{l_{max}} \end{aligned} \tag{3.48}$$

Finally, the normalised residual frame is rounded to the nearest integer and discretised to 8-bits such that:

$$res_{frame} = l_{res_{norm}} \times (2^8 - 1) \tag{3.49}$$

The discretised residual frames are finally passed to the codec to form the residual stream.

Auxiliary stream

The compression algorithm produces an auxiliary metadata which is stored in the form of lookup table containing the l_{min} value of the optimal exposure histogram, the minimum and maximum value of y_res which is used on the decoder side for calculation of the optimal exposure for each decoded base frame and expansion (inverse-logarithm) of the corresponding decoded residual frames.

Decoding and merging to HDR

On the decoder side, both the base and residual streams are decoded. The base frames subsequently undergo colour space change from YC_bC_r to *sRGB* which are then linearised and merged with the expanded residual frame y_{res} to form the decoded HDR frames such that:

$$hdr = base_{rgb} + y_{res} \tag{3.50}$$

3.4 HDR video encoding

Any discussion on HDR video encoding will be incomplete without a discussion on video codecs. The purpose of this section is not to introduce the concepts of video encoding but only to focus in specific areas relevant to the topic at hand. All aforementioned HDR video encoding schemes uses state-of-the-art codecs to encode HDR video frames into an video stream. This section briefly discusses the salient principles and parameters relevant to this thesis and also discusses some of the salient features of current state-of-the-art encoders and their shortcomings in encoding HDR video content.

3.4.1 Overview of codecs

The progress of digital video technology has resulted in higher resolution video frames (a sequence of images) captured by modern video camera sensors which are capable of producing frames in High Definition (1280×720 and 1920×1080 pixels) and Ultra High Definition ($3840 pixels \times 2160$ pixels) resolutions with wider colour gamuts at a standard 24/25/30/60 fps. Equation 3.51 exhibits a straightforward calculation of the required transmission cost for one second of uncompressed *1080p* video captured 30 fps at a typical bit depth of 8-bits/pixel/channel i.e. 24 bits/pixel (bpp):

$$3 \times (1920 \times 1080 \times 30 \times 8) = 1.39 Gbps \tag{3.51}$$

Thus, it is quite evident that an increasing amount of digital data is being produced which is impractical for storage or transmission purposes and can cause extremely high computational demands to manage the data. Fortunately, digital video data contains a large amount

of redundancy which can be discarded for lossy compression purposes. However, it is to noted that there exists a trade-off between video fidelity and output size, also known as the *bitrate*.

Overall, a video codec (encoder and decoder) can be defined as an electronic circuit or a software which performs compression and decompression of captured video frames. The compression is typically lossy although advanced state-of-the-art codecs offer a lossless compression mode which endeavours to retain maximal data possible to perform perceptually lossless compression. A codec is typically made up of a series of several modules where each module might contain a set of colour space transforms, lossy compression algorithms, spatial to frequency transformation techniques and motion compensation and prediction techniques in order to encode successive video frames into a bit-stream with redundant information discarded to reduce the overall output size. Typically, video codecs perform the lossy encoding process of the input data in the following five to six steps namely, a) colour space change (optional), b) reduction of resolution, c) motion estimation, d) discrete cosine transform (DCT), e) quantisation and f) entropy encoding.

Colour space change

As previously discussed in Section 2.4, the HVS has lower sensitivity to colour information compared to luminance information. Therefore, separation of the luma and chroma information facilitates reduction in redundancy and improved compression of input data. In case the input video frames are in RGB format, the codec performs a colour space transformation to YC_bC_r (generalised as YUV) thereby separating the luma and chroma information. This also facilitates the second step i.e. resolution reduction.

Reduction of resolution

Taking advantage of the limitations of HVS, the chroma components (channels) U and V are reduced to half the pixels in horizontal direction (4:2:2 sub-sampling) or both in horizontal and vertical directions (4:2:0 sub-sampling) compared to the luma component. The 4:2:2 and 4:2:0 sub-sampling reduces the data volume by 33% and 50%, respectively [Ric11]. Further details are given in Section 3.4.4.

Motion compensation/prediction

A video sequence exhibits high temporal coherency [3]. Motion estimation provides a technique to predict minute changes in successive video frames. This allows the encoder to encode the entire information of single frame (known as the Intra frame) and subsequently encode only the motion vectors which predict the change in successive video frames, thereby

[3] successive frames being very similar with minute differences (except for scene cuts/changes)

allowing to discard most of the redundant information. Motion compensation and prediction is performed using Intra (I), Predicted (P) and Bi-directional predicted (B) frames. Further details are given in Section 3.4.7.

Frequency transformation

The Discrete Cosine Transform (DCT), similar to Fast Fourier Transform (FFT), allows a video frame to be represented in the frequency domain rather than in spatial domain. This allows easier quantisation of data, the primary reason for data loss in the encoding process. However, a DCT is computationally very expensive and its complexity increases by a rate of $(O(N^2))$. Also, the inability of DCT to decompose a broad signal to high and low frequencies simultaneously forces the codec module to divide the spatial data into small pixel blocks such as 8×8 or 16×16 pixels, also known as *macroblocks*, to ease the computational load. Until this point, most of the redundant data has been discarded but no compression has been affected on the data.

Quantisation

The quantisation step is the primary source of data loss in a lossy compression process. The video frame data in frequency domain are divided by a quantisation matrix which takes into account the limitations of HVS. As the HVS is more reactive to low frequencies than high frequencies, they are preserved with finer quantisation while high frequencies undergo coarse quantisation thus reducing the domain significantly. This is mathematically defined as:

$$F_{quant}(U,V) = \frac{F(U,V)}{Q(U,V)} \tag{3.52}$$

where $F_{quant}(U,V), F(U,V)$ and $Q(U,V)$ are the quantised frequencies, original frequencies and quantisation matrix for the U and V channels, respectively. The quantisation matrix can be varied and controlled by a flag known as the Quantisation Parameter (QP) in order to change the amount of required compression thereby varying the output file size (output bitrate). The usage of QP values is explained later in Section 3.4.5.

Entropy encoding

Entropy encoding is the final step in the encoding process and is performed by two steps; a) Run Length Encoding (RLE) [Pou87] and b) Huffman coding [H*52] which are essentially lossless compression techniques used to compress the data further by an additional factor of three to four [Ric11]. The resultant output data from the entropy encoding process is known as the bit-stream and the output file size determines the bitrate of the video stream.

Codecs: a (very) brief history

The most widely used video codecs are the standards MPEG-1, MPEG-2 and H.264/MPEG-4 AVC (advanced video codec) and the latest H.265/HEVC (high efficiency video codec). The MPEG-1 (formally known as ISO/IEC-11172) standard [LG91] was introduced in 1992 with the aim of providing VHS quality video with the bandwidth of 1.5 Mbps which facilitated video playback from a CD-ROM. It was designed to have forward and backward seeking capabilities along with synchronisation between audio and video. MPEG-2 (formally known as ISO/IEC-13818) [BBQ*97] was introduced in 1994 and facilitated higher fidelity video with slightly higher bandwidth. It was designed to be compatible with MPEG-1 and later on used for DVD and HDTV encoding and decoding. The playback frame rate was locked to either 25 or 30 fps. MPEG-2 was more scalable than MPEG-1 and able to play the same video in different resolutions and frame rates.

The MPEG-4-Part 1 (formally known as ISO/IEC-14496) [Koe02] standard introduced in 1998 was a major development from MPEG-2 which facilitated the production of high-fidelity video with lower bitrates with the primary goal of being used in interactive environments such as multimedia and video communication environments. It also offers re-usability of contents and better copyright protection. While the MPEG-4-Part 1 standard provided many features for multimedia and video communication purposes, the increasing number of services which used high-fidelity video content such as the popularity of HDTV, transmission of media over cable modem, xDSL and UMTS networks require high coding efficiency. In 2001, the Video Coding Experts Group and MPEG *ISO/IEC JTC 1/SC 29/WG* formed a joint video team (JVT) to finalise the draft of a new video coding standard, the H.264/AVC [WSBL03] also known as MPEG-4-Part 10. Similar to previous generation codecs, only the decoder is standarised by imposing restrictions on bitstream and syntax such any decoder implementation conforming to the standard will produce similar output. This limited scope facilitated maximal freedom to optimise the implementation of the encoding and decoding process specific to applications; balancing compression quality, implementation cost and computational load. The improved H.264/MPEG-4 AVC codec, introduced in 2003, offered up to 50% better coding efficiency that the previous generation MPEG-2 [OSS*12]. It introduced a hybrid spatial-temporal prediction model which included variable macro-block structure, sub macro-block motion estimation, motion vectors over picture boundaries, weighted prediction, directional spatial prediction, context-adaptive entropy encoding [4]. Further details of these underlying concepts are explained in [Ric11].

Although the state-of-the-art H.264/MPEG-4 AVC is ubiquitously used in almost all existing digital video applications, the increasing diversity of digital video services and the popularity of Ultra High Definition (UHD) TV and formats such as $4k \times 2k$ and $8k \times 4k$, trig-

[4] Entropy encoding in H.264/AVC is performed using either Context Adaptive Binary Arithmetic Coding (CABAC) or Context Adaptive Variable Length Coding (CAVLC)

gered the need for better encoding efficiency for high resolution videos along with support for stereo or multi-view capture and display. Furthermore, the popularity of mobile devices such as smart-phones and tables along with the increasing number of video-on-demand services impose severe challenges on current networks. The solution to this issue was the introduction of the High Efficiency Video Codec, formally known as the H.265/HEVC standard [SOHW12], first adopted in 2013 and able to provide almost 50% better coding efficiency at similar frame quality levels compared to the H.264/AVC standard. The design structure of HEVC is similar to the H.264/AVC standard albeit with certain improvements. It enhances the motion prediction module of the H.264/AVC by introducing Coding Tree Structures, also known as Coding, Prediction and Transforms units (CU/PU/TU) and 35 directional mode for Intra prediction. The coding and transform units support larger macro-block structures. Whereas the H.264/AVC is fixed up to 16×16 macro-block units, the HEVC macro-block CU can be chosen from a range of 8×8 to 64×64. Compared to the 60 fps encoding limit, the HEVC can encode videos at up to 300 fps. Currently three profiles are supported but a draft of additional five profiles are under consideration with 13 levels per profile. However, the HEVC is upto 300% computationally expensive compared to H.264/AVC, largely due to larger coding units and expensive motion estimation. Comparative evaluations of compression performance between the H.264/AVC and HEVC standard are given in [OSS*12, PDAN12].

The work presented in this thesis uses codec implementations conforming to both the H.264/AVC and HEVC standard as highlighted later in Chapters 6 and 7, respectively.

3.4.2 Colour spaces in video encoding

Video codecs by default, support encoding of video frames in *luma-chroma* formats. in YC_bC_r colour space. A coded picture represents either coded frames or a single field with chroma components C_b and C_r sampled and aligned horizontally with every Y (luma) sample. The H.264/AVC encoder for instance can also take RGB frames as input but eventually converts them into YC_bC_r before encoding. The input format is thus controlled by a flag.

3.4.3 Input file formats

Frames to be encoded using state-of-the-art codecs can typically stored as intermediate file formats such as *.yuv* or *.y4m*. These formats requires the frames to be stored in *luma-chroma* formats and in planar orientation (successive channels) or interspersed format where each row of individual channels are stored in groups. As previously mentioned, if the frames are stored in sRGB format, the same should be passed as an argument to the codec. Also, the resolution of chroma channels can be either reduced while creating the *.yuv/.y4m* files to match one of the chroma sub-sampling formats mentioned later in Section 3.4.4 or the channels can be stored in full resolution and correspondingly mentioned in the codec exe-

cution argument list in case the desired output sub-sampling format is different to that of the input.

3.4.4 Chroma sub-sampling

Figure 3.16: Chroma sub-sampling formats

The HVS has lower acuity for colour differences than for luminance information. Therefore, chroma sub-sampling is an intelligent technique for sampling YUV frames such that the number of colour samples are less or equal (according to requirements) than luma samples. State-of-the-art H.264/AVCs support four sampling formats viz. 4:0:0 (monochrome), 4:2:0 (default), 4:2:2 and 4:4:4. The most popular format is the 4:2:0 sampling, used for all commercial video applications while 4:2:2 sampling is used for high quality colour reproduction. The 4:4:4 sampling format ensures maximum quality albeit at the cost of significantly increased storage requirements. The three formats are explained as below:

4:0:0 subsampling

State-of-the-art codecs such as the reference H.264/AVCs [AMT] support a pure monochrome sub-sampling format where only the luma channel is sampled and the chroma channels are omitted. This ensures that only 4-bytes are required to represent a 4 pixel macroblock. The stream created is much smaller and many of the *backwards* compatible video compression algorithms described in Section 3.3 takes advantage of this subsampling format to create the monochrome detail streams.

4:2:0 sub-sampling

In this sampling format the two chroma channels C_b and C_r are subsampled to half the vertical and horizontal resolution of the Y channel. The $4:2:0$ sampling only requires six samples i.e. four samples for the Y channel and one each for C_b and C_r requiring a total of 48 bits $\equiv$ six bytes to represent a four pixel macroblock.

4:2:2 sub-sampling

In this sampling format the two chroma channels C_b and C_r are subsampled to the same vertical resolution but half the horizontal resolution of the Y channel. The $4:2:2$ sampling requires eight samples i.e four samples for the Y channel and two each for the C_b and C_r requiring a total of 64 $\equiv$ eight bytes to represent a four pixel macroblock.

4:4:4 sub-sampling

In this sampling format the two chroma channels C_b and C_r have the same horizontal and vertical resolution as that of the Y channel. The $4:4:4$ sampling requires 12 samples i.e four samples for the Y channel and four each for the C_b and C_r requiring a total of 96 bits $\equiv$ 12 bytes to represent a four pixel macroblock. A visual description of the above mentioned subsampling formats is give in Figure 3.16.

3.4.5 Bitrate (Output file size)

Bitrate can be defined as the number of bits per second that are transmitted along a telecommunications network and is directly proportional to the output file size of the bitstream. In video compression, the output bitrate is also directly proportional to the quality of the transmitted video stream. Therefore, the bitrate of a video stream can be directly controlled by the determining the required quality of the reconstructed video. As previously mentioned in Section 3.4.1, the quality of the reconstructed video can be controlled by the quantisation parameters (QP values) of a codec. Typically (as in the case of H.264/AVC), the QP values range from 0 to 51 [AMT], where smaller QP values represent better reconstruction quality and therefore larger output file size. QP $= 0$, essentially represents lossless encoding whereas QP $= 51$ represents highly lossy compression with blocking artefacts. The variation of video reconstruction quality with different QP values has been shown later in Chapters 6 and 7 in the form of Rate Distortion (RD) graphs.

3.4.6 Bit-depth (Luma and Chroma)

Bit-depth can be define as the number of bits/pixel/channel allocated to the luma and chroma channels. State-of-the-art codecs such as JM H.264/AVC [AMT] have extended bit-depth support for luma and chroma channels where $n \in [8, 14]$ bits/pixel/channel.

3.4.7 Types of Frames and GOP structure

In H.264/AVC encoding, there are two types of frames available are broadly classified as *intra* and *inter*. Intra frames are known as 'I' frames and inter frames are of two types 'P' and 'B' frames. They are defined as below:

Intra (I) frame coding

The term intra (I) frame coding refers to the fact that the various lossless and lossy compression techniques are performed relative to information that is contained only within the current frame, and not relative to any other frame in the video sequence. Therefore, no temporal processing is performed outside of the current picture or frame.

Inter frame coding

Intra (I) frames do not compress well since coding techniques process video signals on a spatial basis, relative to the information within the current video frame [Ric11]. Therefore, more compression efficiency can be obtained if the inherent temporal or time-based redundancies, are exploited as well.

P frame (predictive coded frame): contains motion-compensated difference (using *Mean Absolute Difference (MAD)* or *Mean Square Error (MSE)*) information relative to previously decoded frames and each predictively coded region within the P frame refers to only one previously decoded frame as the reference frame.

B frame (bi-predictive coded frame): commonly referred to as bi-directional interpolated prediction frames, because the motion prediction can be predicted or interpolated from an earlier and/or later frame. Note, that the previous and subsequent frame can be either intra (I) and/or inter (P or B) frame. The quintessential advantage of the usage of B frames is coding efficiency. The size of a B-frame is $\approx 25\%$ when compared to an I-frame.

Group of Pictures (GOP): The GOP is a group of successive pictures within a coded video stream. The GOP structure, specifies the arrangement of intra- and inter-frames in a coded video sequence. Note that each video sequence consists of successive GOPs where each GOP always starts with an I-frame. There are no strict rules for GOP structure but based on the video coding efficiency expected from the encoder. If achieving good compression ratio is the primary objective then the number of the B-frames in between 'I' and 'P' frames, need to be increased. However, if image (frame) quality is the primary objective then the length of the GOP needs to be attenuated, frequency of 'P' frames need to be increased. The default reference H.264/AVC GOP is *"I-B-B-B-P"*.

3.4.8 Codec implementations

The reference implementation of the H.264/AVC [WSBL03] and HEVC [SOHW12] are reference versions only and is primarily used for research purposes and as a proof of concept. These implementations are unable to use the capabilities of modern multicore CPUs and do not include computational optimisation such as multithreading and CPU optimisations such as the usage of MMX, SSE2 and SSSE2 registers. Therefore, for practical usability purposes, optimised versions of these reference implementations have been built for research and commercial purposes. These include the *x264* codec [Orga] which is an optimised version of the H.264/AVC standard. The profiles suppported by *x264* are baseline, main and high. The codec also supports various presets from *ultrafast* to *veryslow* which controls the encoding speed. However, the compression efficiency (output file size) is inversely proportional to the speed of compression. Also, unlike the reference H.264/AVC implementation which supports upto 14-bits/pixel/channel, the maximum bitdepth supported by x264 is 10-bits/pixel/channel.

Similarly, the reference HM-H.265 [SOHW12] provides a reference implementation of the HEVC standard supporting upto 16-bits/pixel/channel encoding. The faster implementation of the HEVC standard suitable for most research and commercial purposes is the x265 [Orgb] which is able to take advantage of multicore CPUs, MMX, SSE2 and SSSE2 registers. Futhermore, the profiles and presets supported are similar to the x264 implementation. However, it is to be noted that x265, similar to the reference HM-H.265 is an ongoing project and is subject to continuous improvement.

Finally, the discussion on codec implementaions can be concluded with a brief discussion on *ffmpeg* [Bel]. It is the largest video and audio encoding library available to date which provides intrinsic support for most video and audio encoding formats including legacy formats such MPEG-2. A very detailed user documentation of *ffmpeg* is available from [Bel].

3.5 Summary

The goal of this chapter was not to introduce the reader to the vast literature available on HDR video compression but to provide a brief introduction to some of the fundamental concepts of the HDR video compression. To that end, the two main approches to HDR video pre/post processing algorithms have been discussed in Section 3.1 along with their advantages and shortcomings. The reader was then introduced to the fundamental concepts of transfer functions relevant to HDR video compression only as the concepts of transfer function based HDR video compression have been extensively used later in Chapter 7. Next, the reader was introduced to several existing state-of-the-art HDR video compression algorithms including some which are under active development. Later, in Chapters 6 and 7, some of these algorithms have been comprehensively evaluated by means of objective

and subjective evaluation techniques to determine the various design decisions taken in each algorithm and to evaluate the performance of each algorithm thereby investigating the advantages and shortcomings of each. Finally, the user was introduced to some of the fundamental concepts of video codecs since they are essential for a complete understanding HDR video compression process on the whole.

Chapter 4

Evaluation

THIS chapter introduces the reader to a variety of objective and subjective evaluation techniques typically used to test the quality of HDR and LDR videos. These techniques are a part of the Quality of Experience (QoE) studies conducted on image and video quality as available in the literature. The QoE studies include objective evaluation by means of multiple full-reference image/video quality assessment (QA) metrics both for HDR and LDR images/videos. It also includes psychophysical studies by means of rating-, ranking- and pairwise-comparison-based subjective evaluations. Finally, this chapter provides a brief overview of the previous evaluations conducted on TMO and HDR video compression which serve as the basis of the works discussed later in Chapters 5, 6 and 7.

4.1 Objective Quality Assessment

This section provides a brief overview of some of the objective QA metrics used to evaluate the quality of HDR and LDR images primarily for compression purposes. Although QA metrics can be classified as *no-reference*, *single-reference* and *full-reference* metrics, only *full-reference* metrics are used for video compression purposes where the quality of the reconstructed images/video frames are compared to that of the reference images/video frames. Although there are many *full-reference* QA metrics available for LDR and HDR image quality evaluation, most LDR metrics cannot be used for HDR image quality evaluation. Therefore, this discussion primarily focuses only on *full-reference* QA metrics which can be used for HDR image/video quality evaluation and used throughout this thesis. These metrics can be broadly classified as:

- *Dynamic range dependent QA metrics:* Mathematical QA metrics which are typically designed for LDR images and catered in accordance with the limitations of LDR displays.
- *Dynamic range independent QA metrics:* Mathematical QA metrics which can be used for both HDR and LDR images/video frames.

- *Structural QA metrics:* QA metrics which typically measure the structural similarity between the reference and distorted images/video frames.
- *Perceptual QA metrics:* QA metrics which typically use an HVS model to predict the quality of images/video frames taking into account the response of the HVS to input stimuli as well the limitations of the visual system.

4.1.1 Dynamic range dependent QA metrics

Dynamic range dependent QA metrics, often also known as mathematically convenient metrics [CH07] have found widespread usage in the image processing community primarily due to their mathematical convenience. These QA metrics typically operate solely on the intensity of the distortions and are dependent on the dynamic range of the target display. The QA metrics are typically used for LDR-LDR image/video frame pair evaluation where the maximum luminance of the target display is limited to $\approx 300\ cd/m^2$ and the image pair to be evaluated are discretized to 8-bits/pixel/channel.

Mean Square Error:

Mean Square Error (MSE) or Root Mean Square Error is one of the most straightforward QA metrics to compute the error between two similar images. These metrics simply compute the mean of the energy difference between the reference image and the distorted image as formulated in equations 4.1 and 4.2.

Let I_{ref} be the reference n-bit (typically 8-bit) image and I_{dis} be the n-bit distorted image. Therefore, the energy of the distortion can be defined as $E = I_{ref} - I_{dis}$ and the MSE between the reference and distorted image is formulated as:

$$MSE = \frac{1}{N}\sum_{i=1}^{N} E_i^2 = \frac{\|E\|}{N} \tag{4.1}$$

where E_i denotes the i^{th} pixel value of E, $\|\cdot\|$ denotes the L_2 norm and N denotes the number of pixel in a 2-D matrix (image channel). Therefore, from 4.1, the RMSE between I_{ref} and I_{dis} can be computed as:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N} E_i^2} \tag{4.2}$$

Peak Signal to Noise Ratio (PSNR)

The Peak Signal to Noise Ratio (PSNR) is a closely related variant of the MSE/RMSE and perhaps the most widely used QA metric for judging image/video frame quality in compression related applications (involving image/video frame distortions). It is defined as

the ratio between the peak value of a signal and the power of the distorting noise that affects the quality of the representation. Therefore, for an *n-bit* signal, PSNR can be defined as:

$$PSNR = 10 \cdot log_{10}\left(\frac{(2^n-1)^2}{\sqrt{MSE}}\right)$$
$$= 20 \cdot log_{10}(2^n-1) - 20 \cdot log_{10}\left(\frac{\|E\|}{N}\right)$$

which in terms of a 2-D matrix (image channel) representation can be defined as:

$$PSNR = 20 \cdot log_{10}\left(\frac{max(I_{ref})}{\sqrt{MSE}}\right) \tag{4.3}$$

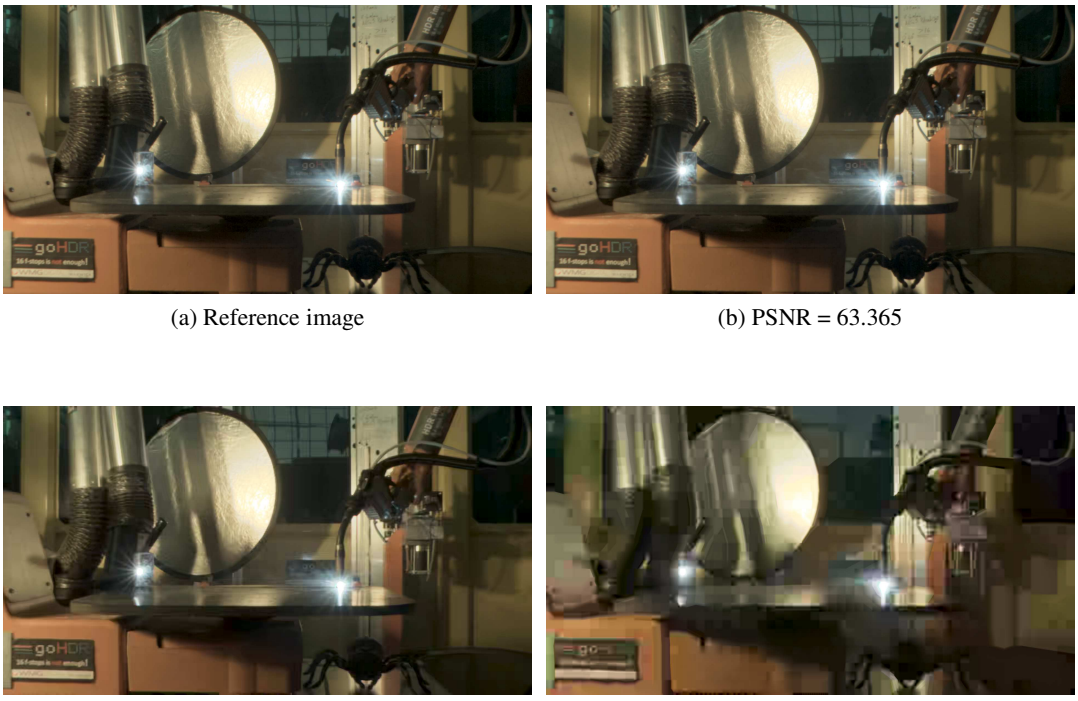
For colour images, say *img*, typically comprised of R,G and B channels, the average PSNR can be computed as:

$$PSNR_{avg} = \frac{1}{N}\sum_{j=1}^{N} 20 \cdot log_{10}\left(\frac{max(img_j)}{\sqrt{MSE_j}}\right) \tag{4.4}$$

where N is the total number of channels and j is the index of the channel under computation. Although PSNR was specifically designed for LDR image/video quality prediction, they can also be used for HDR imaging purposes albeit with a minor modification. To account for the absolute graded (physical) luminance values of HDR images / video frames, a peak luminance L_{peak} needs to be fixed instead of a varying peak power of a signal. In certain cases, it is assumed that the peak luminance of an average scene is $\approx 10^4$ cd/m^2 or in some cases graded to the peak luminance output of the HDR display [SIMa] i.e. 4000 cd/m^2. Therefore, from equation 4.3, PSNR for HDR imaging purposes can be formulated as:

$$PSNR_{avg} = \frac{1}{N}\sum_{j=1}^{N} 20 \cdot log_{10}\left(\frac{L_{peak}}{\sqrt{MSE_j}}\right) \tag{4.5}$$

However, it should be noted that in spite of the widespread usage of PSNR in signal processing and imaging communities, it is widely known that energy-difference QA metrics are poor predictors of image quality and not always analogous to the image quality judged by the HVS since MSE, RMSE and PSNR operate on pixel values, rather than on physical luminance values of the distortions eventually emitted by the display device and perceived by the HVS [CH07]. Later, in Chapter 6, a correlation is established between PSNR (modified for HDR purposes) and the subjective quality as judged by the human eye. A visual description of image quality levels (typically for compression purposes)predicted by PSNR is given in Figure 4.1.



(a) Reference image

(b) PSNR = 63.365

(c) PSNR = 52.422

(d) PSNR = 45.482

Figure 4.1: Examples of predicted image quality using HDR-VQM at different compression quality levels (higher is better).

4.1.2 Dynamic range independent QA metrics

Dynamic Range independent QA metrics are mathematical QA metrics with the added advantage of being independent of the dynamic range of the target display and image pair to be evaluated. These metrics are marginally better in predicting image/video frame quality and also unlike dependent metrics, they can be used for HDR-HDR image pair evaluation. These metrics are typically modifications of LDR QA metrics such as PSNR with a logarithmic extension to accommodate the dynamic range displayed by HDR image pairs and target HDR displays. Typical examples are logarithmic PSNR (logPSNR) and Weber MSE (using Weber Fractions) as derived in equations 4.7 and 4.11.

Let I_{ref} and I_{dis} be the reference and distorted HDR image/video frame pair, respectively. For the sake of convenience, it is assumed that both images are luminance channel only (2-D matrix) where the pixel values are graded in physical luminance values (absolute scale). The first step is to take a logarithm of the image pair such that the pixel values can be mapped to a logarithmic scale. $\mathrm{L_{ref}} = \log_{10}(\mathrm{I_{ref}})$ and $\mathrm{L_{dis}} = \log_{10}(\mathrm{I_{dis}})$

Following the encoding of the pixel values in a logarithmic scale, the RMSE is calculated as:

$$RMSE_{lum} = \sqrt{(L_{ref} - L_{dec})^2} \tag{4.6}$$

Similar to modified PSNR as shown in equation 4.5, logPSNR is adapted for HDR

usage by fixing the peak signal value to $10^4\ cd/m^2$. With a fixed peak assumed, logPSNR is calculated as:

$$logPSNR = 20.log_{10}\left(\frac{log_{10}(10^4)}{RMSE_{lum}}\right) \tag{4.7}$$

Weber MSE

The Weber-Fechner law as described previously in Section 3.2.1 provides an approximate model of the HVS response to input light stimuli. As shown in Section 3.2.1, the digital representation of the Weber's law is a logarithmic function. This introduces biasedness towards the low intensity regions of the image [HBP*15]. Therefore, for image quality assessment purposes, Ameer et al. [AB08] introduced the Weber fraction to calculate the energy-difference between the reference and distorted images which is formulated as:

$$WF = \frac{(I_{ref} - I_{dis})}{(I_{ref} + I_{dis})} \tag{4.8}$$

where WF denotes the Weber fraction, I_{ref} and I_{dis} denotes the reference and distorted images, respectively. However, this does not remove the biasedness towards low intensity values due to the denominator shown in 4.8. Therefore, the law is modified to mitigate this biasedness and the modification can be formulated as:

$$WF_{mod} = \frac{(I_{ref} - I_{dis})}{I_{ref}} \tag{4.9}$$

Since the Weber fraction function is symmetric, a straightforward modification leads to the formulation of Weber-based error as shown in equation 4.10.

$$W_{error} = \frac{(I_{ref} - I_{dis})}{max(I_{ref}, 1 - I_{ref})} \tag{4.10}$$

The Weber-based error can then be used to derive the Weber-based absolute MSE as shown in equation 4.11.

$$WMSE = \frac{1}{N}\sum_i \frac{\|I_{ref} - I_{dis}\|}{max(I_{ref}, 1 - I_{ref})} \tag{4.11}$$

4.1.3 Structural QA metrics

The HVS being highly sensitive to structural information in the input stimulus [WBSS04] and extracts a significant amount of structural detail from the scene. Therefore, loss of structural information correlates to perceptual loss of quality [WB06]. A number of QA models, therefore, have been proposed which quantitatively measures the loss of structural quality that maybe attributed to the introduction of noise, compression artefacts, pre- and post-processing. In this section, the reader is introduced to some of the most significant and widely used QA metrics which defines the structural similarity between the reference and

distorted images.

SSIM:

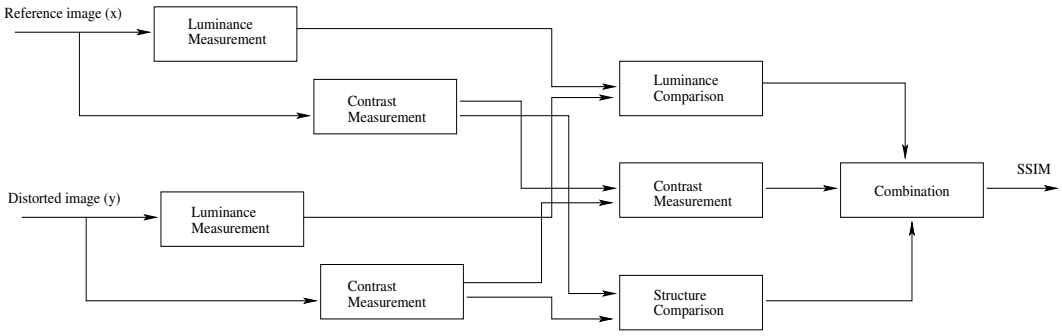


Figure 4.2: Schematic diagram of the SSIM measurement system

Wang et al. [WBSS04] introduced a full-reference QA metric known as the Structural Similarity Index Metric (SSIM) which is an extension of the original Universal Quality Index (UQI) [WB02]. The proposed QA metric estimates the loss in image quality by estimating the loss in luminance, contrast and structure from patches drawn from the same location of the reference and distorted image pair. This can be formulated as follows:

Let $x = \{x_i | i \in [1,N]\}$ and $y = \{y_i | i \in [1,N]\}$ be the two patches from the reference and distorted images, respectively where i is a positive integer. Therefore, the loss in luminance $l(x,y)$, contrast $c(x,y)$ and structure $s(x,y)$ is formulated as in equation 4.12.

$$\begin{aligned} l(x,y) &= \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \\ c(x,y) &= \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \\ s(x,y) &= \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3} \end{aligned} \tag{4.12}$$

where C_1, C_2 and C_3 are small constants included to prevent divide-by-zero conditions. $\mu_x, \mu_y, \sigma_x, \sigma_y, \sigma_{xy}$ are the means, variances and covariance of x and y, respectively. The patch used to compute the loss is an 11×11 circular-symmetric Gaussian function with weights $w = \{w_i | i = 1,2,3,.....,N\}$ with a standard deviation of 1.5 samples, normalised to sum to unity such that $\sum_{i=1}^{N} w_i = 1$. Finally, by taking into account all of the above three factors, the SSIM index is computed as shown in equation 4.13.

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{4.13}$$

A schematic diagram of the SSIM measurement system is given in Figure 4.2.

MS-SSIM:

The multi-scale SSIM [WSB03] was proposed as an extension to the SSIM QA metric [WBSS04] in order to evaluate the image quality between the reference and distorted image pair at multiple resolutions. In this QA metric, the original signals (images) x and y (as previously discussed) are iteratively filtered using a low-pass filter and subsequently down-sampled by a factor of 2. The image pair at full resolution is indexed as 1 and every subsequent iteration is indexed such that index $i \in [2, M]$ where M (typically 5) denotes the number of iterations. The contrast $c_i(x,y)$ and structural loss $s_i(x,y)$ is measured at every iteration whereas the luminance loss $l_M(x,y)$ is measured only at the M^{th} scale. Finally, the combined structural similarity is measured over scales as shown in equation 4.14

$$SSIM(x,y) = [l_M(x,y)]^{\alpha_M} \prod_{i=1}^{M} [c_i(x,y)]^{\beta_i} \cdot [s_i(x,y)]^{\gamma_i} \tag{4.14}$$

In this case, the exponents $\alpha_M, \beta_i, \gamma_i$ are selected such that $\alpha_M = \beta_i = \gamma_i$ and $\sum_{i=1}^{M} \gamma_i = 1$.

4.1.4 Perceptual QA metrics

Objective QA metrics which take into account the response of the HVS to external stimuli into account are known as perceptual QA metrics. These metrics can either be HDR specific perceptual extensions to more widely used energy-difference (see Section 4.1.1 and 4.1.2) and structural metrics (see Section 4.1.3) originally designed for LDR image/video quality purposes or can also be dedicated HDR image/video specific metrics. This section provides a brief overview of four such metrics which are subsquently used later in Chapters 6 and 7.

puPSNR/puSSIM:

Perceptually Uniform (PU) PSNR/SSIM introduced by Aydin et al. [AMS08a] is an extension of the more commonly used quality metrics such as PSNR and SSIM. The authors argue that most LDR metrics input gamma corrected reference and decoded images and assume that pixel values are scaled to be perceptually uniform. This assumption, although valid for darker displays such as the older generations of CRT and LCD (typically with peak luminance of $\approx 80-100 \text{cd/m}^2$ displays are invalid for the much brighter HDR displays with a peak luminance of ≈ 4000 cd/m^2. Therefore, the authors propose a straightforward extension to LDR metrics in order to objectively evaluate the quality of HDR-HDR image pairs without affecting the evaluation capability of legacy LDR-LDR image pairs.

The proposed extension maps physical luminance values within the range of $Y \in (10^{-5}, 10^{9}] \text{cd/m}^2$ to perceptually uniform (PU) code values using a PTF (for details, see Section 3.2.1) and is stored in the form of a look-up-table. The PU encoding is derived from the contrast sensitivity function (CSF) (originally proposed in Daly's Visible Differ-

ence Predictor (VDP) [Dal92]) which predicts the contrast detection threshold for a large range of physical luminance values. The PU encoding additionally ensures backward compatibility with sRGB non-linearity up to $80-100\text{cd}/\text{m}^2$. The contrast detection thresholds of the HVS at varying background luminance values can be mapped using a Contrast vs. Intensity (cvi) curve which can be formulated as:

$$cvi(L,L_a) = (\max_{\mathbf{x}}[CSF(L_a,x)MA(|L-L_a|])^{-1} \tag{4.15}$$

where the CSF is the contrast sensitivity function, x denotes the parameters such as spatial frequency, stimuli size etc., L_a denotes the adaptation luminance and L denotes the background luminance. The function $\text{MA}(\cdot)$ denotes the mal-adaptation. Using the *cvi* function in equation 4.15, the detection thresholds can be formulated as:

$$t(L) = cvi(L, max(L, L_{a-min})) \tag{4.16}$$

where L is the background luminance, L_a is the adaptation luminance and $L_{a-\text{min}}$ is the minimum adaptation luminance that the HVS can detect. Using the threshold estimation function as described the equation 4.16, the forward mapping function to uniformly encode luminance to luma code values can be described as in the recursive equation 4.17

$$f_i = f_{i-1}(1+t(f_{i-1})) \text{ where } f: L' \longmapsto L, i \in [2,3...,N] \tag{4.17}$$

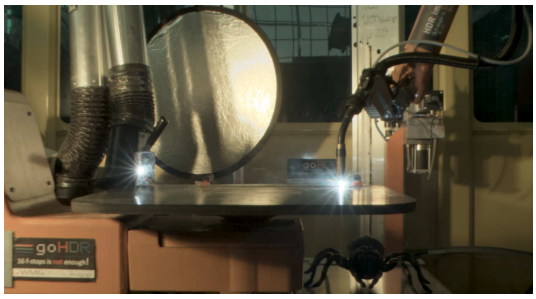
where f_1 is the minimum encoded luminance i.e. 10^{-5} cd/m^2 and N is selected such that f_N is larger than the maximum luminance to be encoded i.e. 10^9 cd/m^2.

The proposed extension can be applied to both traditional energy difference metrics such as *PSNR* or structural similarity metrics such as *SSIM* [SB06]. The proposed QA metric(s) have been used later in Chapters 6 and 7 and it can seen that the image quality prediction of puPSNR/puSSIM has a significant correlation with subjective evaluation results. A visual description of image quality levels (typically for compression purposes) predicted by puPSNR and puSSIM is given in Figure 4.3.

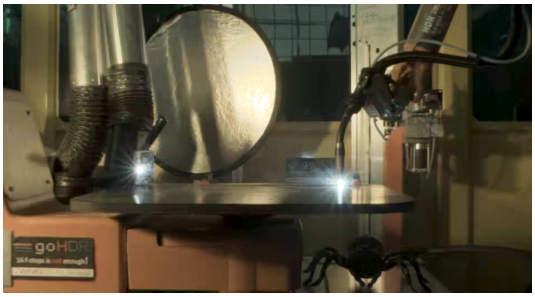
HDR-VDP:

The HDR-Visible Difference Predictor (VDP) proposed by Mantuk et al. [MMS04] is a dedicated HDR extension to the original VDP, a perceptual QA metric proposed by Scott Daly [Dal92]. The proposed metric takes the HVS perception to input light stimuli into account and uses several perception based models such as *amplitude compression, contrast detection, cortex transform* and *visual masking* to predict the quality of HDR-HDR image pairs. Although, the modules in the proposed metric are common to the original metric proposed by Daly, several changes had to be implemented for HDR quality evaluation purposes. Figure 4.4 provides a visual description of the data-flow in the original VDP which

(a) Reference image



(b) puPSNR = 61.327, puSSIM = 0.999



(c) puPSNR = 44.1621, puSSIM = 0.973

(d) puPSNR = 35.287, puSSIM = 0.868

Figure 4.3: Examples of puPSNR and puSSIM predicted image quality different compression quality levels (higher is better).

is required to describe the corresponding changes implemented to modify the original QA metric for HDR-HDR image pairs. The primary change between VDP and the proposed

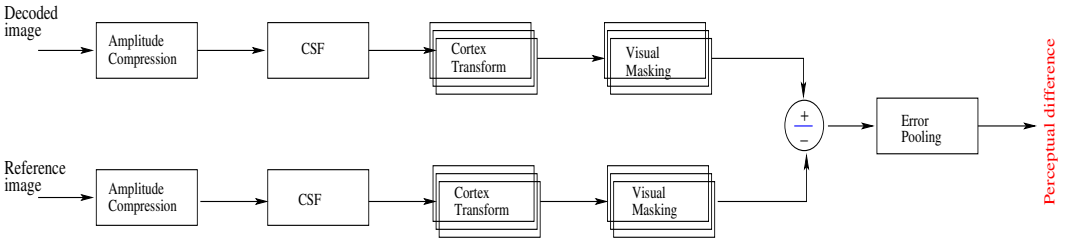


Figure 4.4: Data-flow diagram of the Visible Difference Predictor.

extension HDR-VDP is that the former assumes a global adaptation to luminance whereas the latter assumes that viewers can adapt to every single pixel of the target image. This assumption leads to a conservative estimate of the image quality as well as introducing more reliability to the quality prediction of the proposed metric. The first two stages i.e. *amplitude compression* and *CSF* accounts for the non-linear response of the HVS to input stimuli and the next two stages i.e. *cortex transform* and *visual masking* decomposes the image into spatial and orientation channels to predict perceivable differences. Finally, the probabilities of the detectable changes are taken into account to generate a probability detection map.

The proposed metric does not change the later stages and modifications are mainly

restricted to the first two stages. Unlike the VDP which models the photo-receptor response to input stimuli, the proposed metric uses a perceptually uniform JND scale to non-linearly transform input physical luminance values to perceptually uniform code values. This transformation is analogous to the steps explained previously in Section 4.1.4 derived from the concepts described in Section 3.2.1. A more detailed derivation of the *t.v.i* function is given in [MMS06]. Also, unlike the original VDP where the CSF was responsible for modelling the loss of sensitivity and normalisation of contrast to JND units, the CSF model in HDR-VDP does not require any normalisation since luminance values are already scaled to JND units. The CSF to model the loss in sensitivity in JND units can be formulated as:

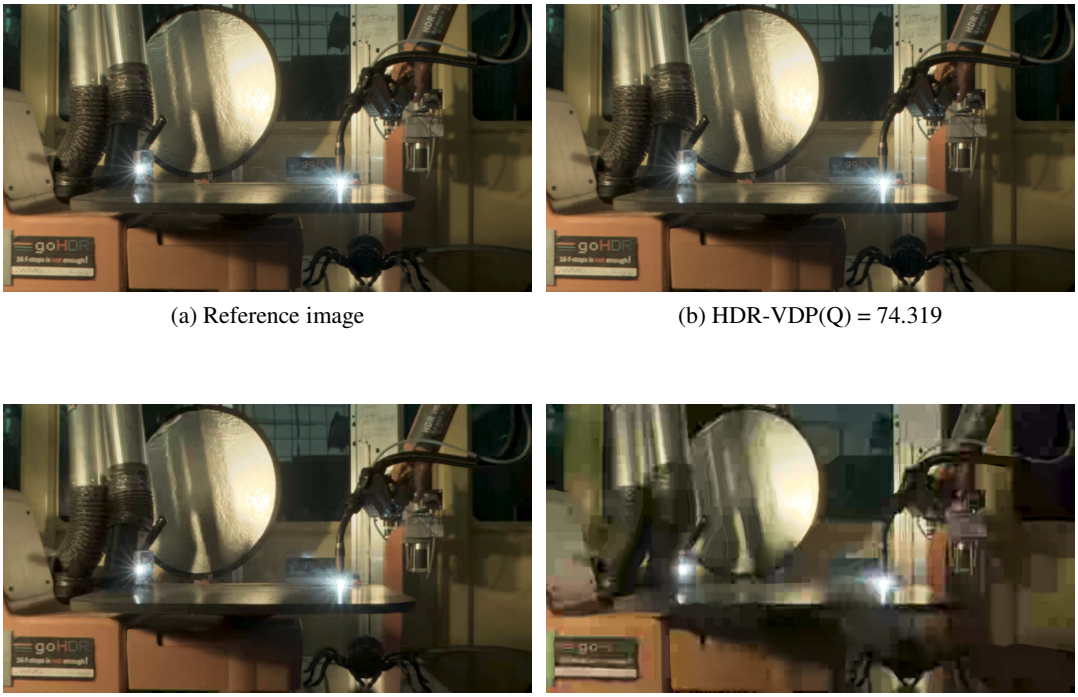
$$CSF_{norm}(\rho, y_{adapt}) = \frac{CSF(\rho, y_{adapt})}{max_{\rho} CSF(\rho, y_{adapt})} \tag{4.18}$$

However, in case of HDR images, a single CSF is insufficient to model the HVS adaptation response. Therefore, the authors filter the image multiple times in frequency domain using a CSF for different adaptation luminance. Subsequently, the images are converted back to spatial domain and pixel values are linearly interpolated. In order to account for scotopic, mesopic and photopic vision the filtration process is repeated for adaptation luminance values $y_{adapt} \in \{10^{-4}, 10^{-3}, 10^{-2}, ..., 10^{3}\}$ cd/m^2. Results suggested that the proposed metric performed better than the original VDP for high luminance regions as the original metric is unable to predict visible differences in high luminance regions.

The HDR-VDP proposed in [MMS04] was subsequently overhauled in HDR-VDP-2 [MKRH11] where the primary contributions were a) generalisation of a broad range of viewing conditions, b) the proposal of a comprehensive visual model derived from a comprehensive psychophysical evaluation which takes into account several properties of the HVS such as intra-ocular light scatter, photo-receptor spectral sensitivity, separate rod and cone pathways, intra- and inter-channel contrast masking and spatial integration and finally, c) improvement of supra-theshold (i.e. distortions clearly visible to the human eye) quality metric predictions. However, an issue with HDR-VDP-2 was that the error prediction was accomplished by the pooling of errors in several frequency bands where the pooling weights were determined by optimising an existing LDR dataset which limits the accuracy of prediction in high contrast scenarios. Secondly, the optimisation was performed on a relatively small set of images and was unconstrained which led to negative pooling weights which were not easy to interpret. Therefore, Narwaria et al. [NMDSLC15] proposed an extension to HDR-VDP-2 which addressed the issues with error pooling. The pooling weights were re-optimised on a combined dataset of LDR and HDR images which resulted in more effective prediction for both LDR and HDR test conditions. Secondly, the optimisation was formulated to be constrained such that the resultant weights can be computed in a bounded manner.

The work presented in this thesis uses the latest optimised version of the HDR-VDP

and unless otherwise stated, HDR-VDP mentioned henceforth refers to the latest optimised version HDR-VDP-2.2 [NMDSLC15]. A visual description of HDR-VDP2.2 predicted image quality levels is given in Figure 4.5.



(a) Reference image

(b) HDR-VDP(Q) = 74.319

(c) HDR-VDP(Q) = 45.5106

(d) HDR-VDP(Q) = 35.2072

Figure 4.5: Examples of HDR-VDP2.2 predicted image quality at different compression quality levels (higher is better).

HDR-VQM:

Narwaria et al. [NSC15] proposed a dedicated HDR video QA metric based on signal processing, transformation and frequency based decomposition of HDR video frames. The proposed metric avoids computationally expensive motion analysis and video quality is measured based on spatio-temporal analysis. The metric was tested with 90 HDR video sequences and found to be a better predictor of video quality than some of the existing HDR QA metrics. The proposed metric takes into account that HDR signal values are generally proportional to the physical scene luminance but not equal to it. Furthermore, unlike LDR pixel values, the concept of fixed upper bound or white point does not exist in case HDR pixel values. Therefore, in the absence of standardisation and presence of inherent limitations of HDR displays, HDR reference and decoded signals have to pre-processed to match the peak luminance capacities of existing HDR displays.

The proposed metric performs pre-processing step where native HDR signals (labeled N_{hdr}) are first linearly scaled to match the HDR display capabilities. The display

processed HDR signals (D_{hdr}) are then considered to contain emitted luminance. These emitted luminance signals are then transferred to perceived luminance where the emitted luminance is converted to JND scaled luma code values using the PU encoding as described earlier in Section 4.1.4 which takes into account the perceptual non-linearity of the HVS. Subsequently, the reference and decoded video frames are analysed using spatio-temporal analysis to create an error video which contains the localised perceptual errors between the reference and decoded video frames. This is performed using the log-Gabor filters to compute the perceptual errors at different scales and orientations. The filters are used in frequency domain and can be defined using polar co-ordinates such that:

$$H(f,\theta) = H_f \times H_\theta \tag{4.19}$$

where H_f and H_θ are the radial and angular components respectively. The filtration process can be formulated as:

$$H_{s,o}(f,\theta) = exp\left(-\frac{log(\frac{f}{f_s})^2}{2\cdot(log(\frac{\sigma_s}{f_s})^2)}\right) \times exp\left(-\frac{(\theta-\theta_o)^2}{2\sigma_o^2}\right) \tag{4.20}$$

where $H_{s,o}$ is the filter denoted by spatial scale index s and orientation index o, f_s denotes the normalised centre frequency of the particular scale, θ is the orientation, σ_s is the radial bandwidth, θ_o defines the centre orientation of the filter and σ_o is the angular bandwidth denoted by $\Delta\Omega = 2\cdot\sigma_o\sqrt{2\log(2)}$. The reference and decoded video frames are decomposed into sub-bands by multiplying the frames in frequency domain using the filter defined in equation 4.20 and subsequently converted into spatial domain by an inverse DFT operation. Subsequently, the errors in the resultant sub-bands are computed at different scales and orientations. The errors from the spatio-temporal neighbourhoods correspondingly undergo short- and long-term spatio-temporal pooling to obtain a global video quality score. Figure 4.6 provides a visual description of the overall quality assessment pipeline of HDR-VQM.

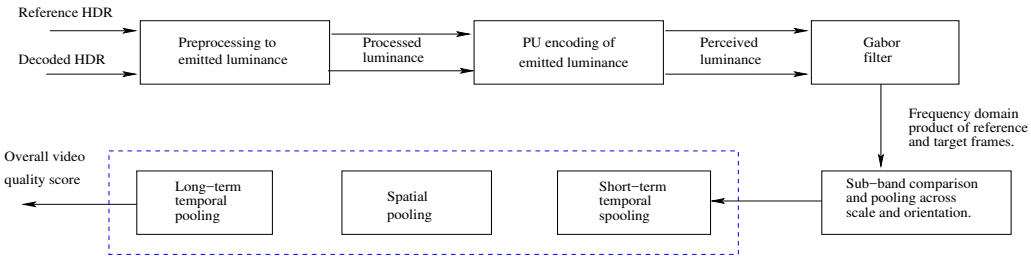
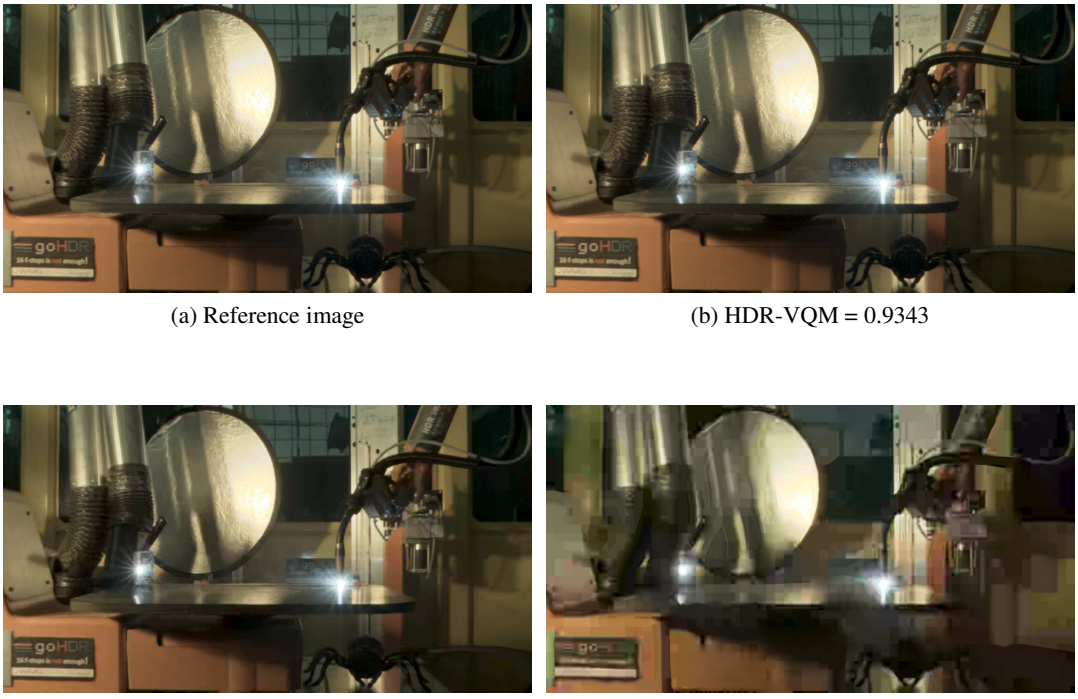


Figure 4.6: Schematic diagram of the HDR-VQM pipeline.

The proposed metric was tested against other full-reference QA metrics including dedicated HDR QA metrics such as HDR-VDP-2.2 and the predictions of these metrics were correlated with subjective evaluations using 25 paid observers. The authors claim that HDR-VQM achieves relatively higher video quality prediction accuracy than other

full-reference QA metrics. Additionally, while other perceptual QA metrics deal with only luminance and is colour blind the second iteration of this QA metrics are able to predict subjective quality taking chroma into account. Unless otherwise stated, the results shown later in Chapters 6 and 7 uses the second iteration of HDR-VQM and the overall perceived quality is judged taking both achromatic and chromatic channel information into account. A visual description of HDR-VQM predicted image quality levels is given in Figure 4.7.



(a) Reference image

(b) HDR-VQM = 0.9343

(c) HDR-VQM = 0.4049

(d) HDR-VQM = 0.1124

Figure 4.7: Examples of HDR-VQM predicted video reconstruction quality at different compression quality levels (higher is better).

4.2 Evaluation of HDR QA metrics

The research and development of a plethora of image and video QA metrics have resulted in a large body of research in the evaluation of such QA metrics. This section introduces the reader to some of the most relevant works in this area.

One of the first QA metric evaluations was conducted by Avcıbaş et al. [ASS02] where the authors used a number of QA metrics categorized into pixel difference, correlation, edge, spectral, context and perception based measures for still image compression applications on various test images distorted by JPEG compression, Gaussian blur and additive noise. Using statistical techniques such as analysis of variance (ANOVA), this work reveals that QA metrics based on spectral magnitude error, HVS absolute norm and edge stability are most suitable for detecting image artefacts such coding error and blur. Sheikh

et al. [SSB06a] conducted an extensive subjective quality assessment using 779 distorted LDR images evaluated by 24 human participants and 10 QA metrics in order to check the consistency of QA metrics. Although not exhaustive, the QA metrics selected in this work represents different classes of QA algorithms. The paper concludes that although multiple QA metrics perform well on multiple image datasets, none of the QA metrics performed at par with subjective quality assessment with 95% confidence interval and further research was required to develop QA algorithms which matches subjective quality assessment. Seshadrinathan et al. [SSBC10] conducted a large scale objective and subjective video quality assessment (VQA) involving several independent state-of-the-art video QA algorithms and 38 human participants who were tasked to assess 150 distorted videos, created from 10 reference videos, using four commonly encountered distortion types. This work concludes that dedicated video QA algorithms such as spatial and temporal versions of Motion Based Video Integrity Evaluation (MOVIE) [SB09] perform significantly better than still-image QA algorithms such as PSNR, VSNR [CH07] and SSIM [WBSS04] Furthermore, subjective and objective results analysed using Spearman's Rho rank correlation and Pearson's correlation tests demonstrate that spatial and temporal MOVIE has higher correlation with subjective evaluation. However, it is to be noted that the mentioned evaluations were all conducted on LDR image and video datasets using QA metrics specifically designed for LDR image and video content.

In comparison, substantially less research has been conducted on the evaluation of dedicated HDR QA metrics such as puPSNR, puSSIM, HDR-VDP-2.2, DRI-VQM and HDR-VQM (see Section 4.1 for details) on HDR image and video content. A few evaluations have been conducted to test the performance of the dedicated HDR QA metrics. Čadík et al. [ČAMS11] conducted an evaluation of HDR-VQA metrics with a dataset consisting of six HDR sequences using an HDR display and concluded that although the predictions by DRI-VQM and HDR-VDP are most suited for HDR-HDR image pairs, executing DRI-VQM becomes prohibitively expensive for sequences with greater than VGA resolution. Azimi et al. [ABDD*14] tested the correlation between seven QA metrics and subjective quality scores with a dataset of 40 HDR video sequences and five types of distortions. The work demonstrates that HDR-VDP-2 [MKRH11] outperforms all other QA metrics when measuring compression induced distortions and has the highest correlation with the subjective quality scores. However, VIF [SB06] using PU encoding produces the best overall (tested against all distortions) results. Similar benchmarking evaluations of QA metrics for HDR image/video content have been conducted by Valenzise et al. [VDSLD14], Mantel et al. [MFF14] and Hanhart et al. [HBK*14] and Minoo et al. [MGBL15].

4.3 Subjective Quality Assessment

A significantly large number of computer graphics application typically produce images or videos as output often comparing the quality of the proposed algorithms and reproduction techniques using state-of-the-art objective quality metrics. While objective quality metrics especially full-reference perceptual metrics as described previously are often accurate in their quality prediction and have been shown to correlate well with subjective experiments, their primary restriction is the training set of distortions [SSB06b]. The accuracy of such metrics decreases with the growing variety of distortions [PLZ*09]. Therefore, the final judgment of quality required to convincingly prove the superiority of performance needs to corroborated by user studies with the help of potential users or reviewers. Given the range of distortions that are present in computer graphics applications it is unlikely that user studies will completely be replaced by objective metrics [MTM12]. However, such user studies although more convincing than objective evaluation are typically more tedious and tend to produce noisy results when conducted inappropriately and the interpretation of the results are non-trivial [MTM12]. This section introduces the reader to some of the basic subjective evaluation techniques required to conduct an effective evaluation of image/video content in an increasingly large number of applications. Additionally, this section provides an overview of some of the techniques to design such experiments and conduct appropriate statistical analysis such that the resultant data can be confidently accepted. Some of the techniques described in this section have been used in Chapters 5 and 6 to conduct subjective evaluation of HDR and LDR video content.

Subjective evaluation of image/video content are typically conducted with the help of rating, ranking and pairwise-comparison based experiments. This section describes each technique in brief detail.

4.3.1 Rating based experiments

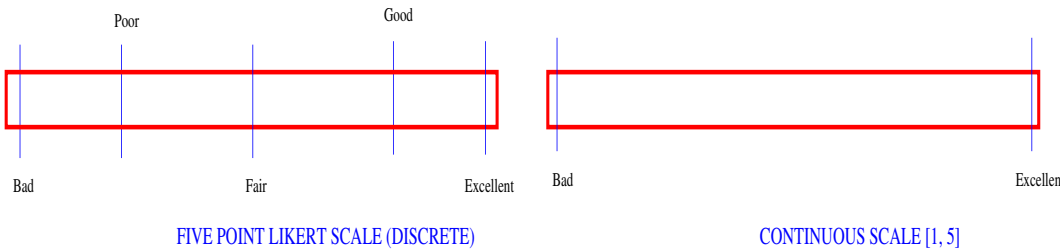


Figure 4.8: Schematic diagram of a likert (discrete) and a continuous scale.

Rating based experiments can be broadly classified into two groups i.e. single and double stimuli categorical rating. In a single stimulus rating an image or video content is displayed for a short duration and the observers are requested to rate the quality of the displayed content on a scale of [1 - N] where higher is better. Typically, in many applica-

tions the rating scale is designed such that the scale $S \in [1,5]$ where the categories are *bad, poor, fair, good* and *excellent* [MTM12] (see Figure 4.8). In some cases, the continuous rating scale is favoured over the five-point Likert-type scale in order to avoid quantisation artefacts [Ass03] (see Figure 4.8). Additionally, such rating techniques also contain a hidden reference (the reference stimuli (image/video sequence) against which other stimuli are tested) randomly presented to the observer to avoid bias.

In a double stimulus rating based experiment, the reference and target image/video content are presented to the observer at the same time typically by means of a dual-display. In case of a single display, the contents are randomly presented one after the other. The primary advantage of rating based experiments is the time required to conduct the experiments. The observers can either execute the task using a controlling GUI or can even mark their rating preferences using a score sheet which can later be digitised for analysis. For a single stimulus experiment, the number of trials required is $n+1$ for n conditions where one extra trial is required for the *hidden* reference.

4.3.2 Ranking based experiments

Ranking based experiments are a more deterministic technique to subjectively evaluate image/video quality. Here the participants are tasked to rank a series to candidate stimuli (image/video content) against a known reference where the basis of ranking is the closeness or resemblance of the candidate stimuli with that of the reference. Similar to the rating based experiments, ranking based experiments, might or might not contain a hidden reference. Also, single/double stimulus ranking based experiments can be conducted where the candidate stimulus is always shown along with the reference stimulus to the participants. Typically, in such an experiment, the candidate stimuli are ordered from $[1-\mathrm{N}]$, where lower is better and N is the number of candidate stimuli.

The primary disadvantage of the ranking based experiments is the time required to conduct such an experiment as the participants have to compare all the candidate stimuli before ordering according to their preference. Ranking can also be indirectly conducted using forced choice pairwise comparisons as explained later in this section. Later, in Chapters 5 and 6, both rating- and ranking-based experiments have been conducted in order to obtain user experience of HDR videos and ranking of HDR video compression algorithms.

4.3.3 Pairwise comparison based experiments

Pairwise comparisons can also be broadly classified into two groups. The first group *ordering by forced choice* requires the participants to choose between a pair of candidate stimuli with similar content but processed with different conditions [MTM12] according to their preference. Observers are forced to choose one candidate in random when they perceive no difference between the candidates. There are several advantages of pairwise comparison

techniques such as fewer problems with the obtained subjective data [BADC11] as compared to rating and the existence of standard statistical techniques to determine the significance of inferred ranks as compared to ranking [BADC11]. However, the main disadvantage of pairwise comparison techniques is the time required to conduct such experiments. Although, there is no time limit, it requires more trials to compare each pair of conditions which can be formulated as $0.5 \times (n \cdot (n-1))$, where n are the number of possible conditions. Although, a full comparison is ideal, the number of trials can be limited using a balanced incomplete block design as described in [MTM12, GT61] or using a sorting algorithm to choose the comparison pairs [SF01].

Although, the forced choice comparison determines of the order to viewing preference, it does not quantify the difference between the stimuli presented. The second group of pairwise comparison techniques are classified as *pairwise similarity judgements* where the participants are not only asked to order the stimuli according to their preference but also to indicate the difference between each pair of stimuli presented, on a continuous scale similar to rating. In case the observer perceives no difference, the marker can be set to '0'. Such experiments are more deterministic and informative albeit at the cost of experiment time. Further details about the comparison methods is available in [MTM12].

4.4 Subjective quality assessment in HDR

During the past decade or so, a considerably large body of research has been conducted on the evaluation of HDR tone-mapping, image and video compression. This body of research can broadly be classified into three groups primarily the subjective evaluation of HDR image/video tone-mapping operators, evaluation of HDR image/video compression algorithms and evaluation of HDR specific QA metrics. Sections 4.4.1 and 4.5 introduce the reader to some of the relevant works conducted on these research areas.

4.4.1 Evaluation of tone-mapping operators (TMOs)

A significant body of research has been conducted on tone mapping techniques to map static HDR images and video sequences to their corresponding LDR versions in order to store and display them using legacy image/video infrastructure. The tone-mapping operators (TMOs), proposed to date, can be classified as global or local TMOs. In addition, they can also be classified as non-temporally coherent TMOs or temporally coherent TMOs suitable for video tone-mapping applications. Furthermore, the availability of a multitude of TMOs has in turn led to the considerable body of research conducted in order to evaluate the TMOs, most of which were conducted by means of subjective experiments in controlled environments using a number of evaluation techniques such as rating, ranking and pairwise comparison.

Drago et al. [DMMS03] was one of the first to conduct a subjective evaluation of

TMOs wherein four different HDR scenes were tone-mapped using seven different TMOs which included the photographic TMO [RSSF02], Tumblin-Rushmier TMO [TR93] and the Retinex TMO [MS06]. The study was conducted by 11 participants by means of a pairwise comparison technique without the *reference* HDR. Results suggested that the Photographic TMO was preferred over other TMOs and this TMO along with Uniform Quantisation and Retinex TMO was described by participants as *better looking images*. This study presented a methodology for measuring the performance of TMO using subjective data [BADC11]. However, the number of participants and HDR data-set was too small to draw any significant conclusions [BADC11].

Ledda et al. [LCTS05] conducted the first TMO evaluation using an HDR *reference*. 48 participants evaluated six different TMOs applied to 23 images using a pair of LDR displays along with an HDR display. The obtained data was analysed using pairwise reference scores [Dav63] along with coefficients of agreement and consistency [Ken48]. The analysed results suggested that the photographic TMO and the iCAM 2002 image appearance model [MFH*02] performed best on the whole. The study presented a robust methodology to evaluate TMOs with a large data-set covering a variety of scenarios and involving a large number of participants.

Kuang et al. [KYJF04] conducted an extensive evaluation of TMOs by studying the viewer's preference and accuracy of TMOs to reproduce real-world scenes. The evaluation was conducted with six TMOs including image appearance models with 33 participants and three different experiments involving pairwise comparison, ranking and rating based psychophysical evaluation techniques. The results obtained from the three experiments suggested that tone-mapped colour and grayscale images were correlated, the viewing preference of users were correlated with TMOs which were able to reproduce better details in shadow areas, preserve the overall contrast and colourfullness and finally the viewing preference of users were highly correlated with the scene reproduction capability of the TMO. Several other TMO evaluations, along similar lines have been conducted such as the ones conducted by Yoshida et al. [YBMS05b], Čadik et al. [ČWNA08], Narwaria et al. [NPDSLC15], Urbano et al. [UMM*10] and Melo et al. [MBDC14].

More recently, Eilertsen et al. [EWMU13] conducted a subjective evaluation where several temporally coherent TMOs were evaluated by means of a pairwise comparison technique. Results demonstrated that several TMOs introduced video artefacts such as flickering, ghosting and redundant saturation. Furthermore, it suggested that relatively less complex global TMOs can outperform complex local TMOs for video application. The work is of particular interest since it evaluates several TMOs for video applications out which one of the temporally coherent TMOs, proposed by Mantiuk et al. [MDK08] has been used in the work described later in Chapter 5.

It is to be noted that the above mentioned TMO evaluations were conducted with the basic assumption that although static HDR images or HDR video sequences are pre-

ferred over a tone-mapped LDR version, they are not compatible with legacy infrastructure. Therefore, the alternative is to evaluate a plethora of TMOs to identify which TMOs are capable of maximal scene reproduction.

An interesting work in this regard was presented by Akyüz et al. [AFR*07] where the authors question the veracity of this fundamental assumption that HDR content is better than LDR content. The authors conducted a series of subjective experiments in order to determine the best technique to display LDR images on state-of-the-art HDR displays and to identify which stages of the HDR pipeline are perceptually most critical. The first experiment conducted as a part of this study used 10 different static HDR images and generated several LDR versions of each. The HDR image was subsequently displayed on a *Brightside DR-37-P* [Tec] HDR display with a peak luminance value of 3000 cd/m^2 and the LDR images were displayed on a commercially available *Dell UltraSharp 2007FP*. Results suggest that although the basic assumption that the HDR image representation would be preferred over LDR holds, it might not necessarily be the case since tone-mapped images have been ranked second to the original HDR representation. Furthermore, the study also determines that although tone-mapped images preserve more details and visibility in general, compared to a single exposure representation of the scene, it might lead to visual unnaturalness in the process as viewers are used to seeing over and under exposed areas in single exposure images. This might lead to a result where tone-mapped images have no statistically significant difference with that of single exposures. The evaluation presents some interesting results and is of particular relevance in this thesis as the work described later in Chapter 5 extends this evaluation for HDR video content.

4.5 Objective and subjective evaluation of HDR video compression algorithms

Despite the extensive research that has been conducted into development and evaluation of QA/VQA metrics for both LDR and HDR content, little has been done to evaluate existing HDR video compression algorithms (as outlined previously in Section 3.3) using both QA metrics and subjective experiments. Koz et al. [KD12] conducted a comparative survey on HDR video compression which compares the two different (non-backward and backward compatible) approaches to HDR video compression as explained earlier in Section 3.1. However, this work has a few shortcomings. First, it does not bring together objective and subjective evaluation techniques to provide a comprehensive evaluation. Second, it focuses on the two approaches to HDR video compression thereby largely ignoring the evaluation of individual algorithms across a large set of sequences.

Recently, Hanhart et al. [HRE15] conducted an evaluation of nine HDR video compression algorithms submitted in response to MPEG CfE [LFH15] to evaluate the feasibility of supporting HDR and WCG content using the HEVC [SOHW12] codec. The paper con-

cludes that the proposals submitted to MPEG can noticeably improve the standard HDR video coding technology and QA metrics such as PSNR-DE1000, HDR-VDP-2 and PSNR-Lx can reliably detect visible difference. However, this work has a few shortcomings. First, the reference sequences are not uncompressed source sequences and are stored as 12-bit non-linearly quantized RGB signal representation. Second, the psychophysical evaluation uses the same training samples as the test samples. Third, the naive participants were instructed to find some specific errors in the video sequences. These issues can result in biased subjective opinions.

Azimi et al. [ABO*15] conducted an objective and subjective evaluation study to compare the compression efficiency of two possible HDR video encoding schemes i.e the PQ algorithm and tone mapping-inverse tone mapping with metadata. The objective evaluation was conducted using four QA metrics and the subjective evaluation was conducted using 18 participants. Results demonstrate the accuracy and monotonicity indexes of the four QA metrics and concludes that the video quality predicted by HDR-VDP and VIF has the highest correlation with subjective results. The correlation was computed using statistical non-parametric tests such Spearman's Rho Rank correlation. Furthermore, it concludes that for specific bitrates, HDR video generated by the PQ scheme were rated higher than the videos reconstructed using the inverse tone-mapping scheme.

Dehkrodi et al. [BDAPN14] conducted a similar evaluation which focuses on the compression efficiency of the HEVC codec compared to the state-of-the-art H.264/AVC codec. The authors use four HDR video sequences and convert them using the PQ algorithm. The converted sequences are then encoded using both the HEVC and H.264/AVC and correspondingly decoded and evaluated using several objective QA metrics. The output sequences are also subjectively evaluated by a rating based experiment using 17 participants. Results suggested that the sequences encoded using the HEVC codec outperforms their H264/AVC counterparts by $\approx 10.18\%$ in terms of quality and yet achieves bitrate savings of approximately $\approx 25.08\%$. Similar evaluations have been conducted by Dong et al. [DNP12], Rerabek et al. [RHKE15], Hanhart et al. [HKE*15], Narwaria et al. [NPDSLC15].

It is to be noted however, that although the works mentioned here present some interesting results, it does not compare the state-of-the-art published or patented algorithms which were proposed before the MPEG CfE call and mostly focuses on the proprietary algorithms presented to MPEG. Furthermore, the works mentioned uses too few HDR video sequences to conclusively draw any generic conclusion. With the growing interest on HDR image/video compression algorithms, it is imperative that a comprehensive objective and subjective evaluation the pre-MPEG algorithms with a robust methodology for evaluation would ideally set the benchmark following which other compression algorithms can be comprehensively evaluation. Moreover, a deep understanding of HDR video compression on the whole can be obtained by such an evaluation. Chapter 6 of this thesis describes such

a work which has been conducted in order to comprehensively evaluate six published and patented video compression algorithms against a large set of HDR video sequences.

4.6 Summary

In this chapter, the reader has been provided a brief overview of both objective and subjective image quality evaluation techniques. The objective evaluation techniques discussed in this chapter include a brief overview of several objective QA metrics representing energy difference, structural and perceptual QA metrics. Later in Chapters 6 and 7, these QA metrics have been used extensively for HDR video compression evaluation purposes. Additionally, this chapter also provides an overview of the previous research conducted on the design and evaluation of LDR/HDR QA metrics, subjective evaluation of TMOs and finally objective and subjective evaluation of HDR video compression algorithms.

In the next chapter, the reader will be introduced to a novel research work which has been conducted to evaluate the viewer's perspective and choice of HDR video over LDR video given certain viewing conditions. The work presented in the next chapter is the first step to answer the research question discussed previously in Chapter 1.

Chapter 5

A Study on User Preference of HDR over LDR Video

5.1 Overview and Motivation

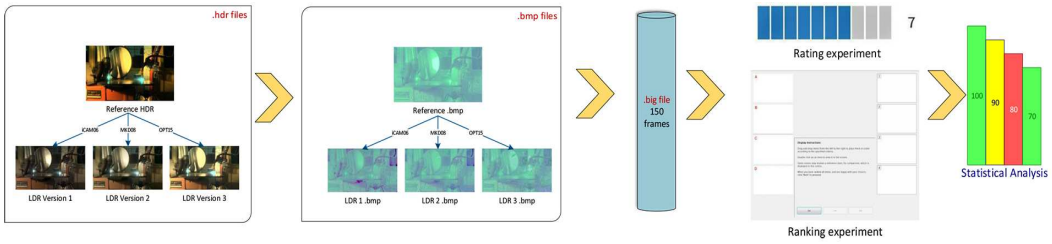


Figure 5.1: An overview of the overall work flow

THE increased interest in HDR video over existing LDR/SDR (standard dynamic range) video during the last decade or so was primarily due to its inherent capability to capture, store and display the full range of real-world lighting visible to the human eye with increased precision, which when compared to the limited dynamic range displayed by LDR video promises to provide a more immersive and realistic viewing experience. Based on this assumption, a large body of research has been conducted in order to process and deliver HDR data by means of image or video tone-mapping and compression algorithms as outlined previously in Chapters 2 and 3, respectively. Although, this assumption is true for most scientific and industrial applications since HDR data provides higher precision than existing 8-bit LDR data, very little has been done to test the veracity of this assumption from an end-users' (viewers') perspective. The work described in this chapter investigates whether HDR video is indeed preferred over LDR video, purely from a viewer's perspective.

A previous related work conducted by Akyüz et al. [AFR*07] on static HDR images suggest that although the basic assumption that the HDR image representation would

be preferred (by end-users) over LDR holds, it might not necessarily be the case since tone-mapped images have been ranked second to the reference HDR representation. Furthermore, the study also determines that although tone-mapped images preserve more details and visibility in general, compared to a single exposure representation of the scene (sequence), it might lead to visual unnaturalness in the process as viewers are used to seeing over and under exposed areas in single exposure images. This anomaly might lead to a result where tone-mapped images have no statistically significant difference with that of single exposure images. Although, the primary research question has been answered in this work, the authors focused on static HDR images only. Furthermore, several advanced perceptually motivated and temporally coherent TMOs have been proposed since. Therefore, the primary motivation of this work is to evaluate whether the findings by Akyüz et al. [AFR*07] hold for HDR video given the current scenario where several perceptual TMOs are able to preserve and reproduce the overall contrast and the tone of the reference HDR scene/sequence. Till date, no such body of work exists for HDR video and a study in order to test the veracity of the basic assumption was the primary motivation of the work presented in this chapter.

In order to test the veracity of this assumption for HDR video content and obtain definitive viewing preference, the requirements were to display the reference HDR video content along with the mapped LDR counterparts on a suitable display such that the full dynamic range of the reference HDR sequence as well as its LDR counterparts can be displayed and conduct one or more subjective experiments such that users are able to provide meaningful quantitative feedback. To that end, six HDR sequences were selected for evaluation purposes. Along with the HDR sequences, three separate HDR to LDR mapping techniques were also selected such that each represent a different class of mapping technique. Using the selected mapping techniques, the sequences were mapped to create three corresponding LDR versions of each sequence. The resultant videos were displayed on an HDR screen where the *reference* HDR representation is absolute luminance graded from 10^{-4} to 4000 $\mathrm{cd/m^2}$ and the corresponding LDR versions are graded from 10^{-4} to 350 $\mathrm{cd/m^2}$. This is done in order to simulate the display capabilities of the HDR display and typical high-end LDR displays, respectively. Subsequently, two subjective studies were conducted by means of a ranking- and a rating-based experiment, to verify the viewing preference of end-users.

The primary contributions of this work are:

1. An indication by means of two subjective experiments that HDR is significantly preferred from mapping methods.

2. Results indicate that the ranking- and rating-based experiments provide similar outcomes which exhibits the preference of HDR over the LDR versions.

5.2 Methodology

This section provides the details of the methodology followed in this work. This includes the choice of HDR to LDR mapping functions, sequence selection, the preparation of materials required for the two experiments and the design and methodology followed to conduct the two subjective experiments. A visual description of the overall work flow is Figure 5.1.

5.2.1 HDR to LDR mapping techniques

Unlike the previous works on tone-mapping evaluation as mentioned earlier in Chapter 4, this work is categorically not a tone-mapping evaluation. Therefore, in this work, three HDR to LDR mapping techniques were chosen such that each represents a different class of HDR to LDR mapping technique and they are as follows:

- A temporally coherent TMO which can also be classified as an SRO (see Section 2.5.6).
- An image appearance model specifically designed for HDR image rendering.
- An alternative technique to extract the optimal exposure from an HDR frame.

The temporally coherent TMO chosen for this work is the Display Adaptive TMO (*mantiuk*) proposed by Mantiuk et al. [MDK08] and the details of this TMO has been described earlier in Section 2.5.6. The primary reason for choosing this HDR to LDR mapping technique is because it endeavours to reproduce the *reference* HDR sequence with minimal visible distortion and also accounts for temporal coherence (for HDR video sequences), ambient lighting and target display. In our case the target display was set to *lcd-bright* in order to exploit the capabilities of the SIM2 HDR display. Also, this TMO in particular performs very well in comparison tests amongst other operators [MBDC14]. A brief overview of this evaluation has been described earlier in Section 4.4.1.

The image appearance model chosen for this work is the iCAM06 HDR image tone compression algorithm proposed by Kuang et al. [KJF07] which is based on the original iCAM framework [MFH*02]. The details of this TMO has been described earlier in Section 2.5.7. The primary reason for choosing this tone compression algorithm is that it provides an HVS based alternative technique to the multitude of available TMOs and yet at the same time predicts and preserves the colourfulness of the original scene. Moreover, unlike the previous iCAM models, this improved model was designed specifically for HDR image rendering.

As opposed to a transfer function based TMO, Debattista et al. [DBRS*15] proposed an alternative exposure extraction technique [HW10] which extracts the optimal exposure from an HDR frame to fit the maximum possible information from the original HDR data within the allowable bit-depth of 8-bits/pixel/channel. Although this exposure extraction technique has been proposed as a part of an HDR video compression algorithm to create

the base LDR stream (the details of which are described in Section 3.3.10), this technique can also be applied in isolation to map HDR image/video content to an LDR image/video frame. The primary motivation behind selecting this mapping technique is that it provides an alternative technique to a myriad of TMOs (choice of which is very application dependent and subjective) to extract the HDR luminance range into a single optimally calculated exposure and maps the exposure into an 8 bit LDR range analogous to an optimally metered 8 bit/pixel/channel image from a camera under varying lighting conditions.

5.2.2 Sequence selection

This section introduces the reader to the HDR video sequences used in this work. Out of a total of 39 HDR video sequences considered, six sequences were shortlisted based on the overall dynamic range of the sequences as well as the source (capture/generation technique) and context of the sequences. A few sequences represented the same scene (same location, same/similar event - different scene cuts) with similar dynamic ranges. In those cases, only one representative sequence was chosen. In other cases, a few sequences were chosen since their overall dynamic range was lower than others (essentially medium dynamic range $approx14-16$-stops) Moreover, it was ensured that the short-listed sequences represent different capture techniques such as the Spheron VR, Arri Alexa, artificially rendered etc.

The shortlisted HDR video sequences (HDRVs), comprising of 150 frames each were so chosen such that the they also represent a wide variety of production techniques. All HDRVs had a resolution of 1920×1080 and were graded (in absolute luminance terms) such that the pixel values are in the range of 10^{-4} to 4000 cd/m^2.

Figure 5.2 and Table 5.1 provides a brief description of each scene along with a tone mapped frame, overall dynamic range and production technique.

5.2.3 Preparation of materials

Following the selection of three HDR to LDR mapping techniques and six HDRVs, three corresponding LDRVs were created for each of the six HDRVs. The output HDRVs and LDRVs (6 HDRVs + 18 LDRVs = 24 in total) produced were in *.hdr* format and in linear RGB colour space. This was necessary since both the HDRVs and LDRVs were subsequently converted to a SIM2 [SIMa] HDR display suitable mode.

Since, the design of the ranking- and rating based experiments required the use of a single HDR display it was necessary to verify the luminance rating of the displayed sequences. The luminance rating of both the HDRV and LDRV frames were verified using the SpectroDuo PR-680 photo-spectrometer [Pho] and it was ensured that the maximum luminance rating of the HDRVs were within 4000 cd/m^2 (catered to be within the range of the SIM2 display) while the luminance rating of the LDRVs were within 350 cd/m^2 (typically representing high-end LDR displays).

(a) Fireplace

(b) Welding

(c) CGRoom

(d) Jaguar

(e) Seine

(f) Tears of Steel

Figure 5.2: Short-listed six HDR video sequences

Name	**Min(Y)**	**Max(Y)**	**DR (stops)**	**Production technique**	**Description**
Fireplace	10^{-5}	4096	25.01	ARRI Alexa	An outdoor winter-night scene with a bright bonfire in the foreground. Scene post processed.
Welding	0.003	5904	19.85	Spheron VR	An indoor scene of a gas welding machine producing intermittent sparks of very high luminance.
CGRoom	0.001	5008	20.82	Rendered	An artificially rendered scene of the dark basement with an overhead lamp swinging as barrels fall from an overhead shelf.
Jaguar	0.0001	4344	25.30	Canon EOS 1Ds Mark III	An side profile indoor shot of a Jaguar E-Type. Bright lights are placed in the room for artificially expanding the scene dynamic range.
Seine	0.005	8864	20.29	ARRI Alexa	Night outdoor scene of the river Seine in Paris with a brightly lit ferry producing the high luminance region of the scene. Scene post-processed.
Tears of Steel	0.017	4088	17.62	N.A.	A clip extracted from the short film produced as a part of the Open Movie project by Blender Foundation.

Table 5.1: Overview of the scenes used for the rating based psychophysical experiment. Here Min(Y) and Max(Y) refers to the average minimum and maximum luminance of the sequence.

Subsequently, both HDRVs and LDRVs were converted to an custom-built intermediate video file format (.big) suitable for displaying the HDR video frames at 30 fps on the

SIM2 HDR display.

5.2.4 Hardware and Software resources

Software resources used for both the ranking and the rating based experiment included the 24 video sequences. Hardware resources included a 47" SIM2 HDR display with a 1920×1080 native resolution, a peak luminance of 4000 cd/m^2 and a contrast ratio of $> 10^6 : 1$ [SIMa]. The LDR display used in the experiments was an Alienware 23" IPS display, also with a 1920×1080 resolution, a peak luminance of 350 cd/m^2 and a maximum contrast ratio of $8 \times 10^5 : 1$. Further the SpectroDuo photo-spectrometer was used for luminance rating verification of both HDRVs and LDRVs.

5.3 Experiment 1: Ranking

This section provides a brief overview of the ranking based subjective experiment which includes a brief discussion about the design of the experiment, materials used, environment of experiment set-up, participant recruitment and the procedure followed in order to conduct the experiment.

5.3.1 Design

The motivation of this experiment was to rank and identify the order of viewing preference of each version (HDR/LDR), across the selected scenes. Based on their judgment of the displayed video quality (overall contrast, brightness, clarity and sharpness), the participants were tasked to rank four versions, which included the hidden *reference* [1], for each of the selected scenes, one at a time. For each sequence they had to view HDRVs/LDRVs at least once. The sequences per scene belonging to each of four versions were randomly presented in order to avoid bias. While ranking the scenes, participants were allowed to view the sequences as many times as required.

The independent variables in this experiment were the selected scenes and the four versions of each scene. The dependent variable in this experiment were the ranks assigned to the four versions for the selected scenes. A within-participants design was employed such that every participant viewed all the scenes.

5.3.2 Materials

For the purpose of the ranking experiment only five HDR video sequences were used as the *Fireplace* sequence was reserved as a demo, results of which, would further be discarded from the final ranking results. This was done since the scene content exhibits high contrast

[1] Hidden reference refers to the reference HDR video sequence which is typically randomly included with LDR versions without the participants knowledge

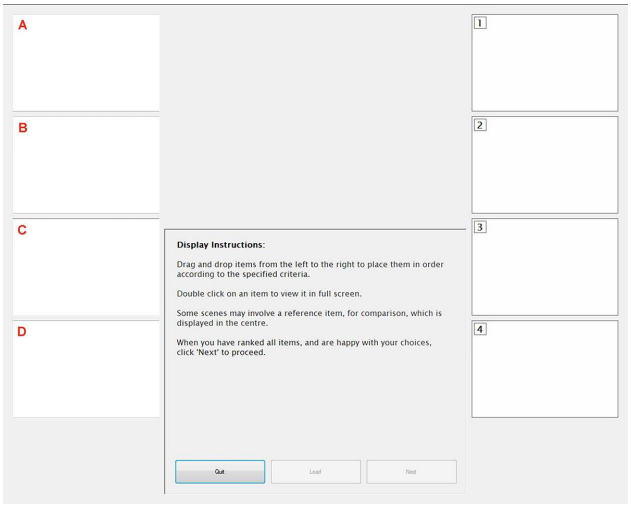


Figure 5.3: Custom GUI used for the ranking experiment to rank the HDR and LDR representations of each sequence based on overall video quality.

(night scene - see Figure 5.2) along with luminance with colour noise which is prone to visual artefacts when mapped to LDR. Therefore, this sequence can be regarded as an ideal training sequence.

Also, a custom graphical user interface (GUI), as shown in Figure 5.3, was specifically built such that it presents four thumbnails each linked to either an HDRV (hidden *reference*) or an LDRV on the left side of the screen. Each thumbnail, when *double-clicked* plays the linked HDRV/LDRV. Participants are tasked to view each of the videos and drag the corresponding thumbnail to the right side of the screen in order of their viewing preference [2]. The instructions for carrying out the experiment are clearly described in the text box in the middle.

5.3.3 Participants

A total of 30 participants took part in this experiment. The total age range was $\approx 25-50$ years. The average age of participants was approximately 28 years and the participants were from various academic (humanities, science and engineering) and corporate backgrounds, typically non-experts in visualisation. All participants had normal or corrected to normal vision and the participants did not have other visual defects. The gender of the participants were not recorded as it was not required due to the anonymous design of the experiment. Necessary ethical approval was obtained under the reference number *PSi*,REGO-2014-1016 and the document is included at the end of this thesis.

[2] In this case preference refers to the overall quality of the video (when viewed by a naïve participant) which includes overall brightness, contrast, noise, colour shifts, flickering and spatial artefacts which is relatively easy to spot.

5.3.4 Environment

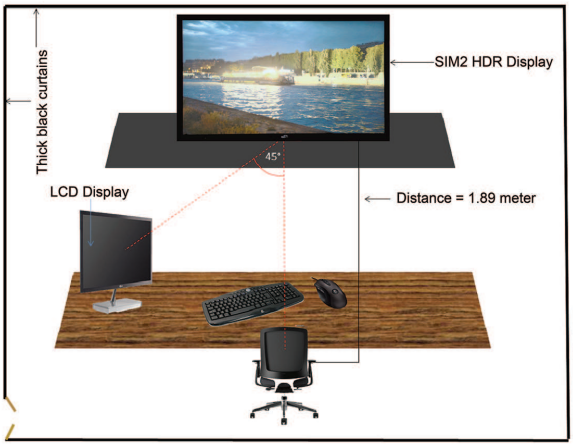


Figure 5.4: Schematic diagram of the ranking experiment setup

Following ITU-R recommendations [ITU12], the experiments were conducted in a room with minimal ambient lighting, below 25 lux, which is within the recommended luminance levels for a typical dark environment [Eng]. The distance between the HDR display and the participant was set to approximately 3.2 times the height of the HDR display (according to the recommendations of ITU-R [ITU12]); at a distance of $\approx 189cm$ with an LCD monitor placed at an angle of 45° (see Figure 5.4). In order to minimise glaring, the brightness and contrast of the LCD monitor was reduced to 25%.

5.3.5 Procedure

The participants were first introduced to the objectives of the experiment which was to judge the overall quality of a video footage (HDRV/LDRV). Along with the verbal introduction, the participants were given a consent form and an information leaflet. Initially, the participants were given a demonstration of the experiment using the demo sequence, the results of which were subsequently discarded from the main results. Upon completion of the demonstration, the participants were asked to proceed with ranking the remaining scenes. Based on their judgement of the displayed video quality, the participants positioned the corresponding thumbnails (labeled [A-D]) to any of the blank positions (labeled [1-4]) by means of the GUI.

5.4 Experiment 2: Rating

This section provides a brief overview of the rating based subjective experiment which includes a brief overview of the design of the experiment, materials used, experiment set up, participant recruitment and the procedure followed respectively.

5.4.1 Design

The independent variables are the six scenes and four versions of each scene. The dependent variable in this case are the scores on a scale of [0-10] given to each of the video sequences by the participants. The participants were tasked to rate four versions, for each of the selected scenes. A within-participants design was used and all the participants viewed all possible combinations of scenes and versions. In order to facilitate the experiment, participants were presented the stimuli in groups of five to eight participants at a time.

5.4.2 Materials

For the purpose of the rating experiment, interactive batch files were created for each group of the participants (see section 5.4.3) such that the total 24 videos (6 HDRVs + 18 LDRVs) are ordered in a random manner to be played sequentially for each group of participants. Furthermore, due to the creation of individual batch files for each group of participants, it was ensured that the ordering of videos for each batch file are also randomised.

5.4.3 Participants

A total of 30 participants were divided into five groups (see Table 5.2). The total age range was $\approx 22-40$ years. The mean age of participants was approximately 25 years and the participants were from various academic (humanities, science and engineering) and corporate backgrounds, typically non-experts in visualisation. All participants had normal or corrected to normal vision and the participants did not have other visual defects. The gender of the participants were not recorded as it was redundant due to the anonymous design of the experiment. Necessary ethical approval was obtained under the reference number *PSi*,REGO-2014-1016 and the document is included at the end of this thesis.

Group number	Number of participants
1	8
2	6
3	6
4	5
5	5

Table 5.2: Detailed breakup of the five groups

5.4.4 Environment

Unlike, the ranking experiment, the rating experiment was conducted in a marginally brighter room. The ambient lighting in the room was below 50 lux, within the recommended luminance levels for a typical dark-dim environment [Eng]. Also, unlike the ranking experiment, where the participant controlled the ranking GUI, the conductor of the experiment

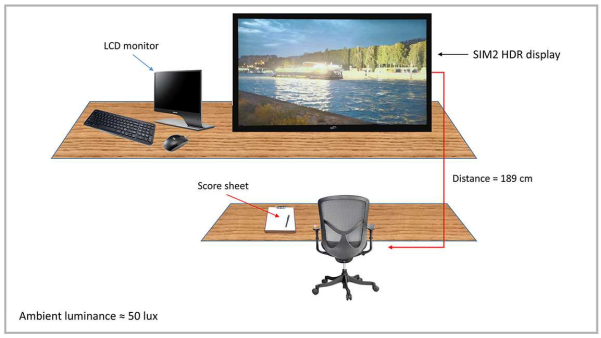


Figure 5.5: Schematic diagram of the rating experiment setup

controlled the interactive batch files for this experiment. Also, the LCD monitor was turned away from the participants during the experiment. A visual description of the rating environment setup is given in Figure 5.5.

5.4.5 Procedure

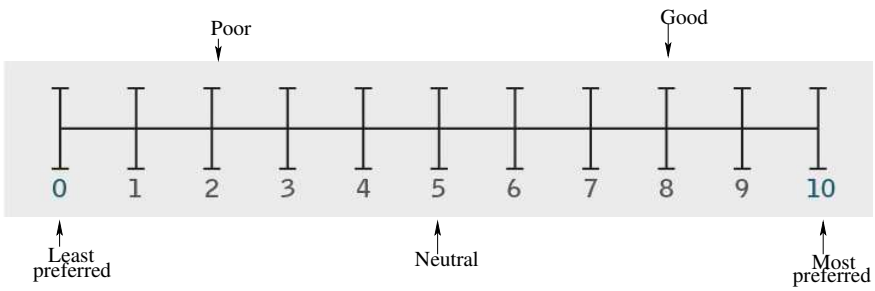


Figure 5.6: Schematic diagram of the rating scale used in the rating experiment such that rate $R \in [0, 10]$, where $R = 0$ denotes least preference and $R = 10$ denotes maximum preference.

The participants were first introduced to the objectives of the experiment and gave their consent for participating. Unlike, the ranking experiment, all six HDR scenes were used in the rating experiment. The participants were tasked to rate the 24 video sequences (played individually) in order of their viewing preference on a scale of [0-10] (see Figure 5.6). However, the participants were also instructed to look for artefacts such as colour shift (common to tone-mapping techniques), flickering (common to non-temporally coherent tone-mapping/compression techniques). The rating was performed on a hard copy score sheet which were later digitised for further analysis.

5.5 Results

This section presents an overview of the results obtained from the ranking- and rating-based experiments and analyses the same.

5.5.1 Ranking results

Let the Null Hypothesis H_0 be that there are so significant differences between the *reference* HDR content and its corresponding LDR versions. The alternate hypothesis H_1 states that there are significant differences between the HDR and LDR versions. The statistical level for analysing the obtained results is assumed to be 0.05 and the sample size (total number of participants) was 30. Furthermore, if H_1 is true, then the Kendall's coefficient of concordance **W** (the degree of mutual agreement amongst participants) can be determined as:

$$W = \frac{2\Sigma}{\binom{N}{2}\binom{t}{2}} - 1 \text{ where } \Sigma = \sum_{i \neq j} \binom{\alpha_{ij}}{2}. \tag{5.1}$$

The significance of W can be analysed using chi-squared statistics such that:

$$\chi^2 = \frac{t(t-1)(1+W(N-1))}{2}. \tag{5.2}$$

χ^2 is asymptotically distributed with $\frac{t(t-1)}{2}$ degrees of freedom, where $t = 4$, represents the number of operators (HDR + 3 LDR) and $N = 30$, represents the number of participants. A significance between scores suggest that the perceived image quality of two operators, when compared with each other are different although no conclusions can be drawn for cases of similarity.

The data obtained from the ranking experiment needs to be tested for homogeneity and any outliers must be removed before further analysis can be performed. To that extent, the data obtained from the ranking experiment is folded across all scenes in order to obtain a grand average. The data is tested for outliers by means of a *histogram* plot and *stem-and-leaf display* method. The outliers are then identified using a *box-and-whisker* plot and are subsequently removed from the raw data. This ensures that the grand average has normally distributed data points.

Following the above mentioned technique, three outliers were identified in the raw ranking data which were subsequently removed, thus reducing the sample size to 27. The resultant data was further analysed using statistical non-parametric tests such as Kendall's coefficient of concordance for K-related samples.

The overall ranking scores demonstrate a significance of $p < 0.05$. Therefore, H_0 is rejected and H_1 is accepted. This means that the ranking results averaged over the sample size of 27 exhibit significant differences between the four operators (versions) for each of the five scenes as well as the grand average of the five scenes. Before the result of the full pairwise comparison on the four operators (on the grand average) is presented, we present the mean ranking scores assigned to each operator per sequence as well as the derived average scores (folded across five scenes) along with their variation denoted by 95% confidence intervals in Figure 5.7.

Next, the results of the full-pairwise comparison on the grand average data is pre-

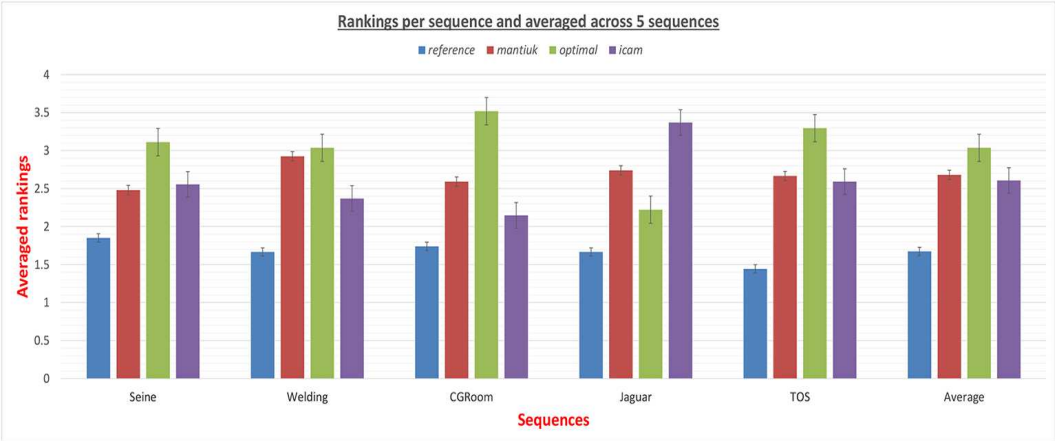


Figure 5.7: Overall ranking scores - per sequence (averaged over 27 participants) and averaged ranks across five scenes(lower is better)

sented in Table 5.3 which demonstrates significant differences between the operators. However, operators within the same group exhibit no statistically significant differences with each other.

Sequence	Methods	Kendall (W)	χ2
Average across 27 participants and 5 sequences	reference 1.54 | mantiuk 2.54 | icam 2.59 | optimal 3.33	0.335	37.137

Table 5.3: Mean ranks with Kendall W, averaged across five scenes and 27 participants (lower is better)

5.5.2 Rating results

Analogous to the process mentioned in Section 5.5.1, the combined results obtained from the rating based psychophysical experiment was folded across the six scenes and the grand average was tested for outliers. Based on the *box-and-whisker* plot, two outliers were identified and removed from the raw data set thus reducing the sample size to 28. Using the resultant data, we present the mean rating scores for each of the four operators per sequence and for the derived grand average in Figure 5.8.

Subsequently, the Null Hypothesis H_0 was tested using the one-way repeated measures Analysis of Variance (ANOVA). The results of the ANOVA indicate a statistically significant difference between the four operators. As the resultant data fails Mauchly's sphericity test, $p < 0.01$, the Greenhouse-Gaussier post-hoc correction was applied, F(1.588, 81) = 10.073, $p < 0.05$, $\eta = 0.272$ which also indicates significant difference between the four operators. Follow up pairwise comparisons, on the grand average indicate, the groups into which the operators can be assigned and the Kendall's coefficient of concordance which denote the degree of agreement amongst the participants as shown in Table 5.4.

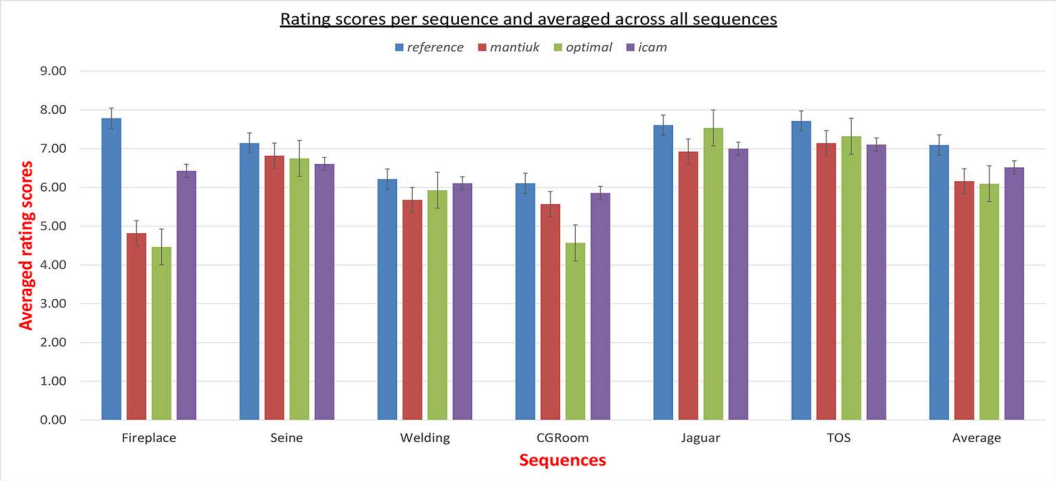


Figure 5.8: Overall rating scores - per sequence (averaged over 28 participants) and average scores across all six scenes and 28 participants (higher is better)

Sequence	Methods	Kendall (W)
Average across 28 participants and 6 sequences	reference 7.10 · icam 6.52 · mantiuk 6.16 · optimal 6.09	0.260

Table 5.4: Mean rating scores with Kendall W, averaged across six scenes and 28 participants (higher is better)

5.6 Discussion

The results from both experiments are overall fairly similar. They indicate a preference for HDR and less of a preference for the LDR mapping methods. There is a distinction between the preference of the mapping methods however, although, for the most part no significant difference between the mapping methods was encountered apart from the *icam* being preferred in the rating experiment.

The mean ranks with 95% confidence interval error bars for each operator as shown in Figure 5.7 clearly exhibit a significant difference between the *reference* HDR and the three LDR versions for each of the five scenes as well as the grand average. However, the difference in-between the LDR versions are less significant. Analysis of the ranking scores, averaged across the five scenes also exhibit the same characteristics wherein the *reference* HDR video exhibit statistically significant difference with that of the LDR versions *mantiuk, icam* and *optimal* as shown in Table 5.4. However, there are no statistically significant differences in-between the three LDR versions. Furthermore, it is to be noted that the Kendall's coefficient of concordance for the grand average ranks, as shown in Table 5.3 exhibit a low concordance value which also indicates a degree of ambivalence amongst the participants.

Similarly, the mean rating scores as shown in Figure 5.8 exhibit a significant differ-

ence between the *reference* HDR and the corresponding three LDR versions for each of the six scenes as well as the grand average. Furthermore, the results of the pairwise comparison from the repeated measures ANOVA demonstrate that the *reference* HDR is significantly different than *icam* which in itself exhibit statistically significant difference with *mantiuk* and *optimal*.

Although the results presented in this work involves HDR video sequences, they bear similarity with the findings of the previous study by Akyuz et al. [AFR*07] which used static HDR and tone-mapped images. Both studies demonstrate a statistically significant difference between the *reference* HDR images/videos and the corresponding tone-mapped versions of the same. Even though many advanced tone-mapping techniques have been proposed since the previous work which endeavours to replicate the overall scene contrast to a higher degree than previous TMOs, some of which has been used in this work, there is evidence that given the correct viewing conditions and properly prepared materials, HDR video supersedes LDR video. However, there are limitations of this study. Only six HDR scenes were used in this work out of which five were used for the ranking based experiments. Results might vary if the number of scenes and HDR to LDR mapping techniques are increased. Furthermore, the viewers were presented with independent visual stimuli which are not a part of any contextual narrative upon which the results might also vary.

5.7 Conclusion

This work asks a fundamental question as to whether HDR video is indeed preferred over legacy LDR video, purely from the viewers' perspective. The technical advantages of HDR video over LDR video and the multitude of TMOs, some of which reproduce a more artistic representation of the original scene were not considered in this work. Therefore, three HDR to LDR mapping techniques were used such that they are able to reproduce the original *reference* to the extent possible and two subjective experiments were conducted with 60 participants in total, 30 in each group (mutually exclusive to each other), both of which demonstrate that given correct viewing conditions, there exists a statistically significant difference between the HDR (more realistic) representation of a scene and its LDR counterparts where the former is preferred by the viewers.

With the end-user preference of HDR video over LDR established and with the pre-notion that HDR video produces significantly large amount of floating point data, it is now necessary to investigate several existing HDR video compression techniques to not only understand the advantages and disadvantages of each but also to establish the best performing HDR video compression algorithm to date by means of a thorough objective and subjective evaluation. Such a work facilitates a deep and clear understanding of the decisions required to design an HDR video compression algorithm and the shortcomings of the existing state-of-the-art as shown in the following chapter.

Chapter 6

Objective and subjective evaluation of HDR video compression

CHAPTER 5 conclusively established the fact that HDR video is preferred over LDR video under the right viewing conditions. However, the primary issue with HDR video is the acute storage and transmission (bandwidth) requirements of HDR file formats due to the floating point values used to accurately capture and store real-world luminance and colour values. To provide a feasible solution, several HDR video compression algorithms have been proposed to date. This chapter introduces the reader to a comprehensive objective and subjective evaluation of the existing state-of-the-art HDR video compression algorithms.

6.1 Overview and contributions

As described in Chapters 2 and 3, HDR video produces a significantly higher amount of data compared to LDR/SDR video. Also, HDR files cannot be directly encoded using legacy or state-of-the-art codecs (see Chapter 3 for details). Therefore, for practical handling of HDR video, a number of HDR video compression algorithms have been proposed to date. Such compression algorithms convert native HDR video data to an encoder suitable format. However, to date, these compression algorithms have been partially compared with each other. The work outlined in this chapter presents a comprehensive objective and subjective evaluation of six previously published and/or patented HDR video compression algorithms and in doing so, follows a detailed and robust methodology for evaluation and qualitative assessment of compressed HDR video content.

The objective evaluation was undertaken using a set of 39 HDR video sequences and seven full-reference QA metrics namely: PSNR, logPSNR, puPSNR, puSSIM, Weber MSE, HDR-VDP and HDR-VQM. The objective QA results are then averaged over 39 sequences at 11 different quality settings to generalise the overall rate-distortion (RD) characteristics

of the compression algorithms. The subjective evaluation was undertaken using six short-listed sequences and two ranking-based subjective experiments with hidden reference at two different quality levels with 32 participants each, who were tasked to rank distorted HDR video compared to an uncompressed version of the same video. Additionally, a correlation was computed between the objective and subjective results for a better understanding of the shortcomings of current objective evaluation techniques for HDR video quality.

Results suggest a strong correlation between the objective and subjective evaluation. Also, non-backward compatible compression algorithms appear to perform better at lower output bit rates than backward compatible algorithms across the settings used in this evaluation.

The primary contributions of this work are:

1. A comprehensive objective evaluation of six HDR video compression algorithms using seven full-reference QA metrics.

2. Two subjective evaluations of the compression algorithms at two different output bitrates using a ranking method with hidden reference conducted with 32 participants each and,

3. An assessment of the correlation between the objective and subjective evaluation results.

6.2 Motivation

As established in Section 4.5, the lack of a comprehensive evaluation of existing HDR video compression algorithms results in an incomplete understanding of HDR video compression on the whole. Furthermore, without a through objective and subjective evaluation it is not possible to understand the design decisions required to create each HDR video compression algorithm and thereby gain a detailed knowledge of advantages and shortcomings of the existing algorithms. Also, as mentioned in Section 4.5, the recent research activity to evaluate compression algorithms is limited to the proposals submitted in reply to the MPEG committee's CfE thereby largely ignoring the previously published/patented HDR video compression algorithms.

The work described in this chapter undertakes a comprehensive objective and subjective evaluation of the previously established compression algorithms which were largely ignored by the MPEG CfE evaluations. The methodology followed to conduct such an evaluation is described later in Section 6.3. The objective and subjective results from this evaluation and corresponding analysis (described in Sections 6.4, 6.5 and 6.6, respectively) not only help identify the best performing algorithm but also provides an in-depth understanding of each compression algorithm along with their advantages and shortcomings thereby providing a basis on which future HDR video compression algorithms can be designed.

6.3 Methodology

This section introduces the methodology followed for the objective and subjective evaluation conducted as a part of this work. It introduces the compression algorithms, sequences, QA metrics used in this work and the overall research method followed for preparing the materials for the objective and subjective evaluation. The individual aspects of the objective and subjective evaluations are presented in Sections 6.4 and 6.5, respectively.

6.3.1 HDR video compression algorithms

For the purpose of this evaluation six HDR video compression algorithms have been selected. This includes most of the published and patented compression algorithms to date as outlined previously in Section 3.3. Out of the compression algorithms selected for this work, two algorithms, *hdrv* and *fraunhofer* follow the *non-backward* compatible approach and four algorithms, *hdrmpeg, hdrjpeg, rate* and *gohdr* follow the *backward* compatible approach. The details of these individual algorithms has been described previously in Section 3.3. The algorithms have been faithfully re-implemented in MATLAB, albeit with minor changes as highlighted below.

hdrv

Mantiuk et al. [MKMS04] was the first dedicated HDR video compression algorithm. Unlike, the original algorithm, the re-implementation (compression and decompression part) are not extensions of the standard MPEG-4 codec. The dedicated modules of the algorithm convert HDR frames to a codec suitable *.yuv* file which is then passed to the codec for encoding and produces a single high-bit depth encoded HDR video stream. On the decompression side the decoded video stream is converted back to HDR frames. The minor changes made in this reimplementation includes the luma bit-depth which was set to 12-bits (instead of 11 in the original work) and the edge-encoding as described in Section 3.3.1 was not implemented as it has been deemed unnecessary due to the advances in video codecs since 2004.

fraunhofer

Garbas and Thoma [GT11] proposed a dedicated *non-backward* compatible algorithm, the details of which were described in Section 3.3.3. This algorithm is a temporally coherent extension of the Adaptive LogLuv transform (see Section 2.3.4) to convert HDR video frames to an codec suitable format. The algorithm was faithfully reimplemented for this work with the one exception of the frame meta-data stored as an auxiliary stream instead of passing them as supplementary enhancement information (SEI) message.

hdrmpeg

Mantiuk et al. [MEMS06] also proposed the first dedicated *backward* compatible HDR video compression algorithm, as previously described in Section 3.3.6. The algorithm was faithfully implemented in MATLAB for the purpose of this evaluation.

hdrjpeg

Ward and Simmons [WS06] proposed a *backward* compatible algorithm to encode HDR images. However, it is quite straightforward to extend this algorithm for video compression purposes. A brief overview of the original still image compression algorithm along with the changes required to extend this algorithm for HDR video compression purposes as used in this work is described in Section 3.3.7.

rate

Lee et al. [LK08] proposed another *backward* compatible algorithm, the details of which are outlined in Section 3.3.8. A block-based extension of this algorithm was later proposed by Lee et al. [LK12] which uses a perceptual quantization similar to *hdrmpeg*. However, this extension is a not a part of this evaluation.

gohdr

goHDR Ltd. [CEB*10] proposed another *backward* compatible algorithm, the details of which are described in Section 3.3.9. This algorithm was faithfully re-implemented in MATLAB for the purpose of this evaluation.

6.3.2 Scene selection

The objective evaluation in this work uses a large set of 39 HDR video sequences with an average dynamic range spanning between $14-23$ stops ($\approx 42-69$ dB). A tone-mapped frame of each sequence along with the overall dynamic range is given in Appendix B and sequences common to both objective and subjective evaluations are marked suitably. Subsequently, based on the overall dynamic range of the 39 sequences, six sequences were carefully short-listed for the subjective evaluation. The selection ensured that the sequences represent a variety of HDR video production techniques, contain a large variation in dynamic range and a large variation in scene content (scenes captured in outdoor, indoor, dark and bright daylight situations). Table 6.1 provides a brief description of the six short-listed sequences shown in Figure 6.1. All sequences used in this work have a full HD resolution of 1920×1080 pixels.

For the purpose of this evaluation the sequences were graded (in absolute luminance terms) such that the pixel values are in the range $\in (10^{-3}, 4000)$ cd/m^2; within the range of

the SIM2 HDR display [SIMa].

(a) Welding (b) CGRoom (c) Jaguar

(d) Seine (e) Tears of Steel (f) Mercedes

Figure 6.1: Short-listed six HDR video sequences used for objective and subjective evaluation.

Name	**Min(Y)**	**Max(Y)**	**DR (stops)**	**Production technique**	**Description**
Welding	0.003	5904	19.85	Spheron VR	An indoor scene of a gas welding machine producing intermittent sparks of very high luminance.
CGRoom	0.001	5008	20.82	Rendered	An artificially rendered scene of the dark basement with an overhead lamp swinging as barrels fall from an overhead shelf.
Jaguar	0.0001	4344	25.30	Canon EOS 1Ds Mark III	An side profile indoor shot of a Jaguar E-Type. Bright lights are placed in the room for artificially expanding the scene dynamic range.
Seine	0.005	8864	20.29	ARRI Alexa	Night outdoor scene of the river Seine in Paris with a brightly lit ferry producing the high luminance region of the scene. Scene post-processed.
Tears of Steel	0.017	4088	17.62	N.A.	A clip extracted from the short film produced as a part of the Open Movie project by Blender Foundation.
Mercedes	0.005	5076	19.688	ARRI Alexa	An outdoor daylight scene of a Mercedes showroom with a parking lot. Scene post processed.

Table 6.1: Overview of the scenes used for the rating based psychophysical experiment. Here Min(Y) and Max(Y) refers to the average minimum and maximum luminance of the sequence.

6.3.3 Quality Assessment (QA) metric selection

Out of a plethora of QA metrics available to measure reconstructed video quality for compression related purposes, seven full-reference QA were used to evaluate the reconstructed HDR video quality in this work. These include dedicated HDR QA/VQA metrics as well as LDR QA metrics extensions. In other words, the QA metrics used for the objective evaluation in this work can broadly be classified into a) extended/modified mathematical QA metrics such as PSNR, logPSNR and Weber MSE [AB08], b) perceptual extensions to mathematical and structural QA metrics such as puPSNR and puSSIM [AMS08a]; and c) dedicated perceptual QA/VQA metrics such as HDR-VDP-2 [NMDSLC15] and HDR-VQM [NSC15]. The details of each QA metric has been described previously in Section 4.1.

6.3.4 Preparations of HDR videos

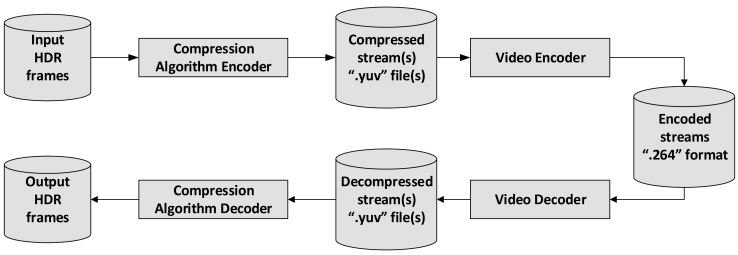


Figure 6.2: Compression Protocol used for the evaluation.

Using each of the six selected algorithms, input HDR frames were converted to an encoder suitable format, creating the intermediate *'.yuv'* files (HDRVs) which were then passed to the video codec (in this case, the H.264/AVC codec [AMT]) creating a raw ('.264') video file. This video file is subsequently decoded and decompressed using the decompression portion of the compression algorithm thus reconstructing the HDR frames. This pipeline allows to plot the rate-distortion (RD) characteristics (quality vs. output bitrate) of each algorithm when the HDRVs were encoded at different quality settings (see Section 6.3.5 for details). Also, all HDRVs were of five seconds duration i.e. 150 frames encoded at 30 frames per second. To preserve the best frame fidelity, the codec sub-sampling format was set to *High 4:4:4* for 3-channel HDRVs and *High 4:0:0* for luma (only) streams, respectively. The Group of Pictures (GOP) structure was set to *I-P-P-P* with an Intra-frame period of 30 frames. Figure 6.2 shows the pipeline.

6.3.5 Quality and bitrate selection

The increase in video quality is directly proportional to the increase in output bits/pixel (bpp). Therefore, it was necessary to evaluate the performance of algorithms at different

quality levels. The output bpp can be directly controlled by setting the quantisation parameter (QP) values of the reference H.264/AVC [AMT] codec where $QP \in [0,51]$ $\forall QP \in Z^{+} \cup \{0\}$, where lower QP refers to a better image quality albeit at higher output bpp.

For the objective evaluation, the HDRVs were encoded at 11 different quality settings such that QP $= 1,5,10,15,20...,50$, where $QP = 1$ represents near lossless compression and maximum output bpp and $QP = 50$ represents a highly lossy compression and minimum output bpp.

For the *backward* compatible algorithms except *rate*, the same QP was set for both the base and residual stream. However, for the *rate* algorithm, the QPs for the residual (ratio) stream were allocated using the Lagrangian optimisation formula mentioned in [LK08], where $QP_{ratio} = 0.77 \times QP_{ldr} + 13.42$ with *automatic rate distortion correction* (a feature of the codec) switched off. Further details about output bpp and bitrate calculations are mentioned in Section 6.3.6.

6.3.6 Bitrate calculation

Let output video file size be f_s and frame resolution be $R_s = frame_{width} \times frame_{height}$ and the number of frames be N.
$\therefore$ for *one-stream* algorithms, the bpp is calculated as:

$$bpp = (\frac{f_{s_1}}{N \times R_s}) \times 8 \tag{6.1}$$

Similarly, for the *two-stream* algorithms, the total bpp is calculated as:

$$bpp = (\frac{fs_1 + fs_2}{N \times R_s}) \times 8 \tag{6.2}$$

the output bitrate is determined as:

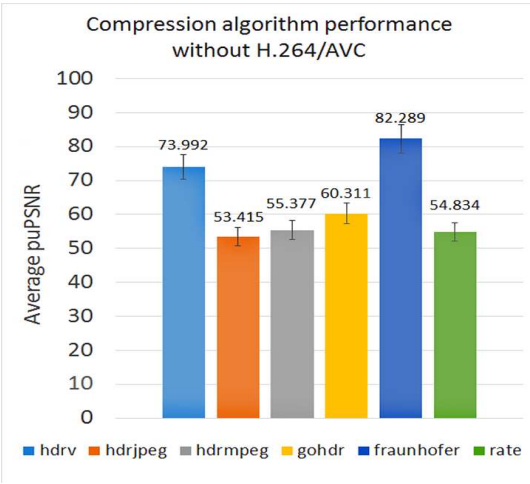
$$bitrate = bpp \times R_s \times frames/sec(fps) \tag{6.3}$$

6.4 Objective evaluation

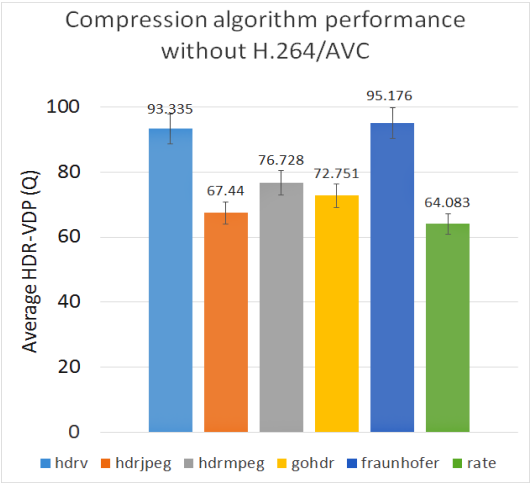
This section demonstrates the results obtained from the objective evaluation following the methodology described in Section 6.3. First, it introduces the coding errors produced by each algorithm followed by a detailed RD performance evaluation of the six algorithms against a set of 39 sequences and finally followed by the RD characteristics of the six algorithms for the short-listed set of six sequences.

6.4.1 Coding errors

Before, going into the performance evaluation of the six compression algorithms, it is important to check the coding errors produced by the algorithms. The coding errors produced by each of the six algorithms can be obtained by following the methodology shown in Figure 6.2, barring the usage of the video codec. Input HDR frames were converted to an encoder suitable HDRV. The HDRV is subsequently decoded using the corresponding decoding function of each algorithm to reconstruct the HDR frames. Video codecs are not used in this work flow. This pipeline tests the maximal reproduction capability of the algorithms without the codec introduced distortions. Since perceptual metrics are designed to predict subjective quality assessment scores, Figure 6.3 shows the coding errors of the six algorithms for the perceptual metrics such as puPSNR and HDR-VDP (averaged across the six short-listed sequences).



(a) Coding errors - puPSNR (higher is better)



(b) HDR-VDP (higher is better)

Figure 6.3: Coding errors of the six compression algorithms averaged across six sequences with 95% confidence interval.

Explanation: It is expected that without the encoder introduced distortions, the reconstruction capability of the compression algorithms should be maximal. However, Figure 6.3 shows that the reproduction capability of *backward* compatible algorithms are significantly lower compared to the *non-backward* compatible counterparts. This anomaly can be attributed to the fact that the *backward* compatible algorithms were designed to take advantage of *dual-loop* encoding scheme (described earlier in Chapter 3), a facility not provided in this pipeline. This can be reaffirmed by the enhanced performance of same algorithms upon the introduction of the codec as shown in Figure 6.4. *Non-backward* compatible algorithms, on the other hand, have no such requirements.

6.4.2 Generalised RD characteristics

This section demonstrates the generalised RD characteristics of the six algorithms upon introduction of the video codec. In this pipeline, the HDRVs from the algorithms for each of the 39 sequences are encoded using the parameters mentioned in Sections 6.3.4 and 6.3.5. Subsequently, the video frames are decoded and reconstructed HDR frames are assessed by the seven full-reference QA metrics. Figure 6.4 shows the full set of results obtained from the seven QA/VQA metrics averaged across 39 sequences. For better clarity, logarithmically scaled plots are used to demonstrate the results.

Although, the RD characteristics presented in Figure 6.4 demonstrate the overall performance of individual algorithms, the results plotted from raw data points do not give a complete perspective. Some of the data points especially for *backward* compatible algorithms are of the order of ≥ 10 bpp which is clearly impractical for storage and transmission requirements. Also, the results were plotted against a large set of HDR video sequences. Therefore, individual algorithms are expected to exhibit variation in both image quality as well as in output bitrate. Figure 6.5 shows the results obtained by fixing output bitrates and interpolating image quality variation across 39 sequences with 95% confidence interval bounds[1] and Figure 6.6 shows the RD characteristics obtained by fixing quality levels and reporting the variation of output bitrates across 39 sequences.

Although Figures 6.4, 6.5 and 6.6 present a generalised set of results for each of the six algorithms, they cannot be directly used to predict the image quality of the six short-listed. Therefore, in Figure 6.7, the RD characteristics of the six short-listed sequences are shown. These results can be directly used to correlate the objective and subjective evaluation results and conduct a combined analysis as discussed later in Section 6.6.

6.4.3 Short-listed RD characteristics

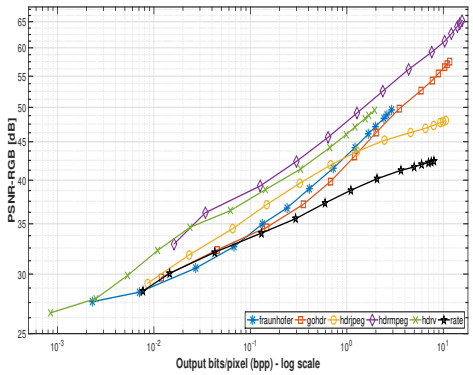
Figure 6.7 shows the RD characteristics plotted from the raw data points for the six short-listed sequences used for the subjective evaluation. Results are presented in a logarithmic scale for clarity.

Next, similar to Figures 6.5 and 6.6, the interpolated set of results for fixed bitrates and fixed quality levels are presented in Figures 6.8 and 6.9, respectively.

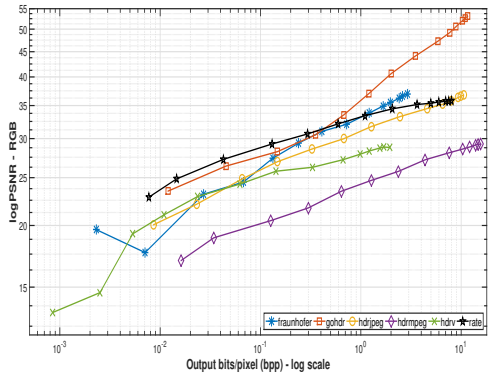
6.4.4 Analysis

This section analyses the results obtained from the generalised RD characteristics as shown in Figures 6.4, 6.5 and 6.6 as well as the RD characteristics obtained from the short-listed sequences as shown in Figures 6.7, 6.8 and 6.9, respectively. A few salient points can be inferred from the objective evaluation results:

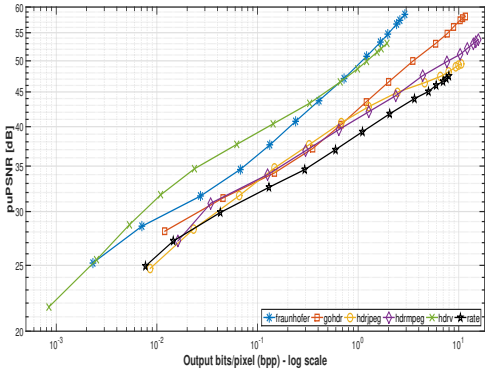
[1] For practical purposes quality variation up to 2.5 bpp is shown. Higher output bitrate (bandwidth) is rarely available in real life.



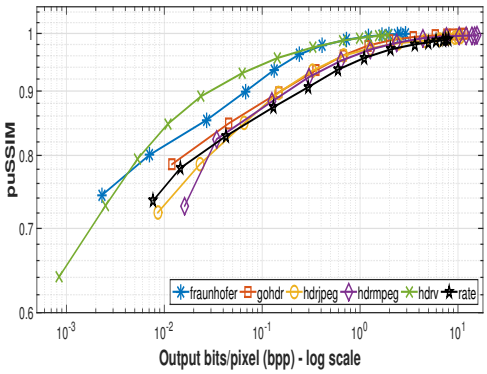
(a) PSNR results (higher PSNR - better quality)



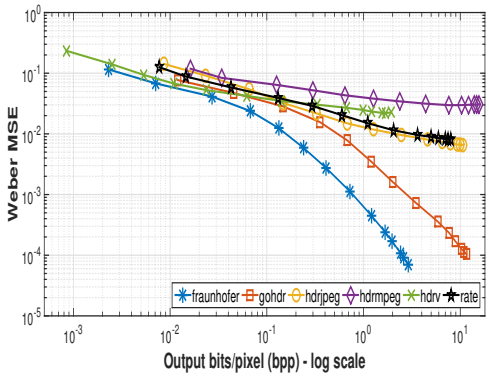
(b) logPSNR results (higher is better)



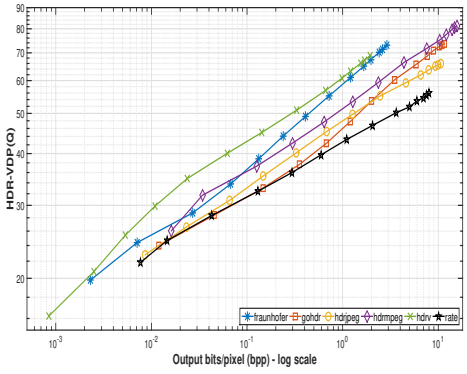
(c) puPSNR results (higher is better)



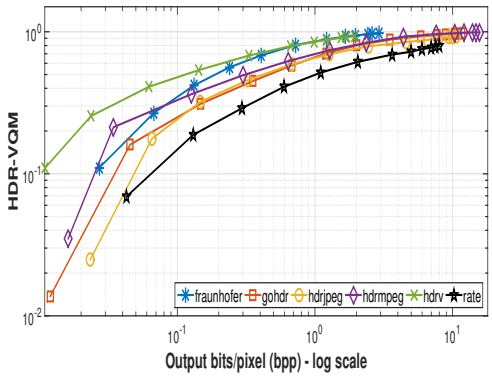
(d) puSSIM results (higher is better)



(e) Weber MSE results (lower is better)

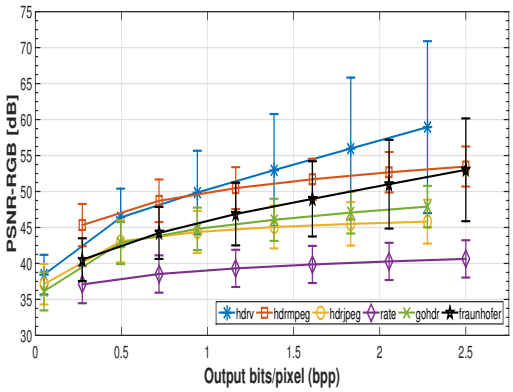


(f) HDR-VDP(Q) results (higher is better)

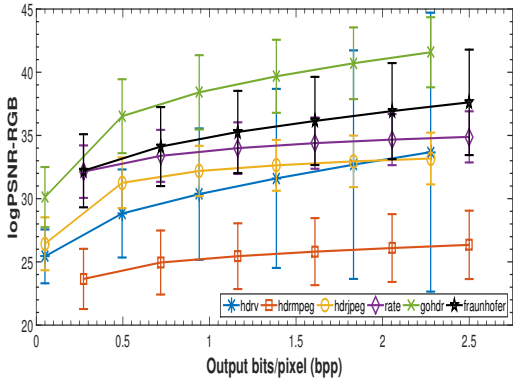


(g) HDR-VQM results (higher is better)

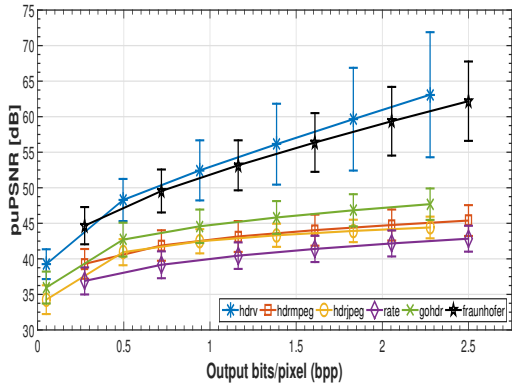
Figure 6.4: Averaged RD characteristics (quality vs output bitrate) of six HDR video compression algorithms against seven QA metrics over 39 sequences. Figures presented in logarithmic scale.



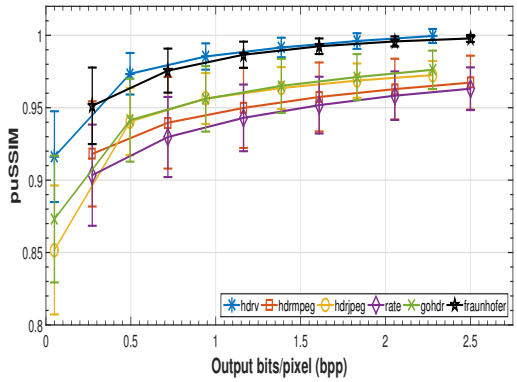
(a) PSNR results (higher is better)



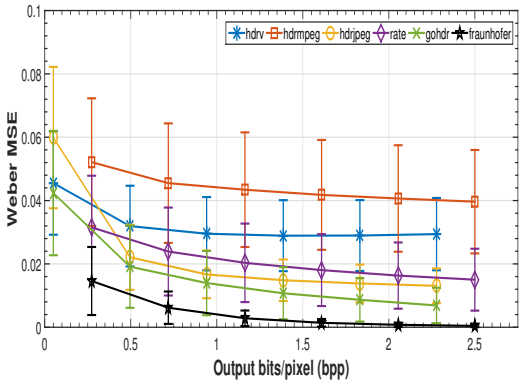
(b) logPSNR results (higher is better)



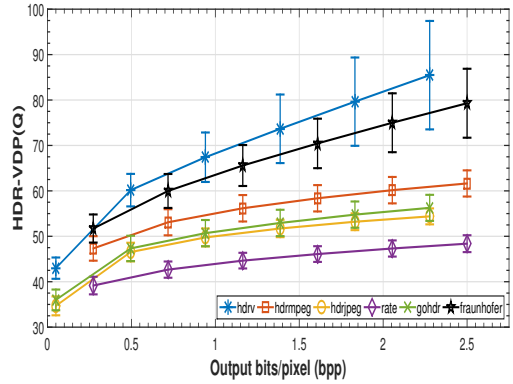
(c) puPSNR results (higher is better)



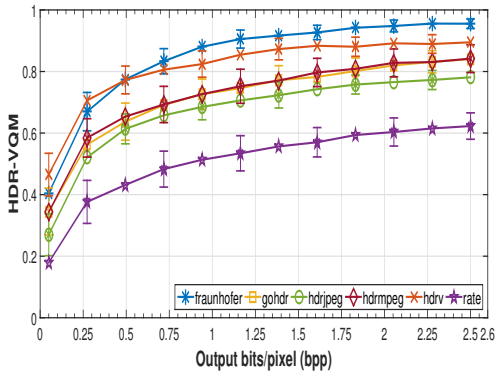
(d) puSSIM results



(e) Weber MSE results (lower is better)

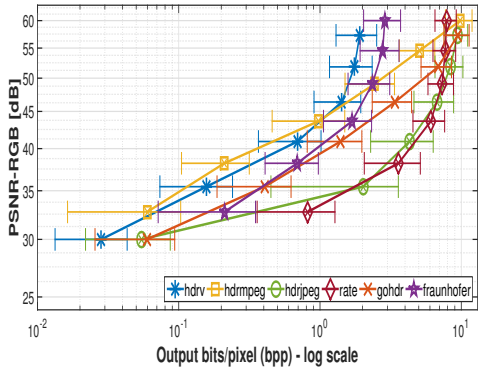


(f) HDR-VDP(Q) results (higher is better)

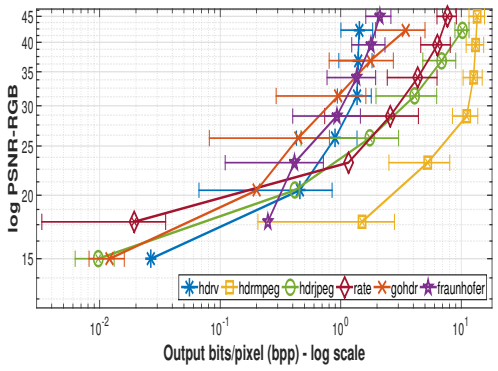


(g) HDR-VQM results (higher is better)

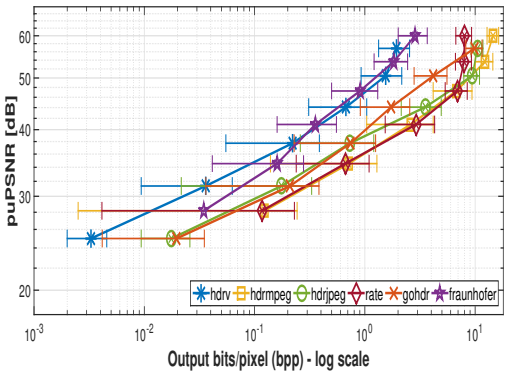
Figure 6.5: RD characteristics - fixed bitrates and interpolated quality levels with 95% confidence interval bounds (presented in linear scale).



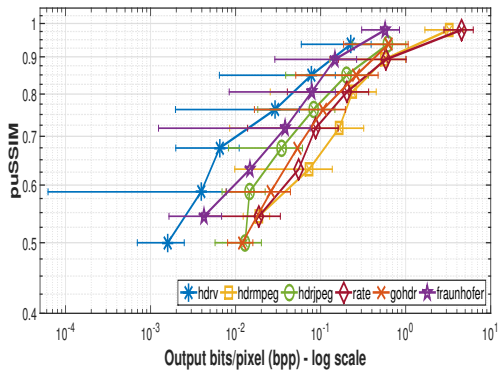
(a) PSNR results (higher is better)



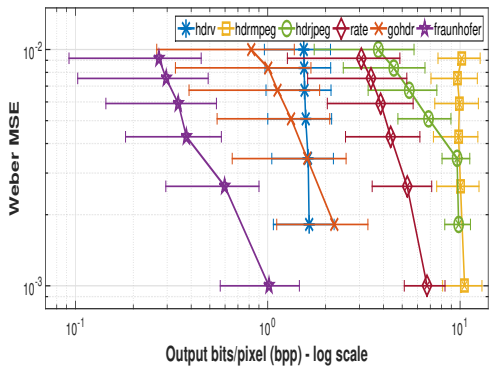
(b) logPSNR results (higher is better)



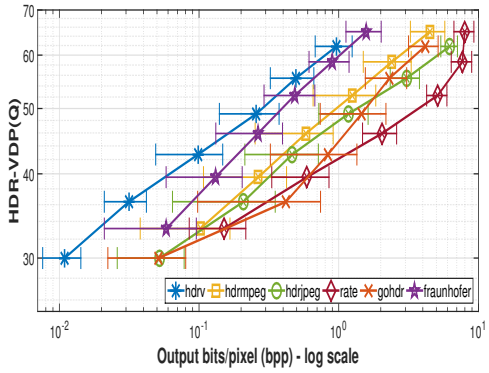
(c) puPSNR results (higher is better)



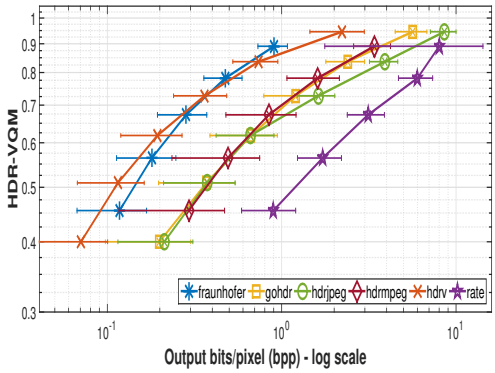
(d) puSSIM results (higher is better)



(e) Weber MSE results (lower is better)

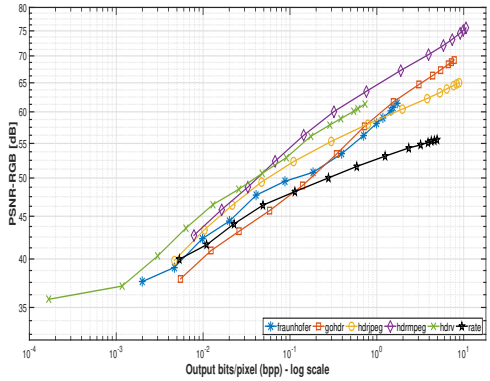


(f) HDR-VDP(Q) results (higher is better)

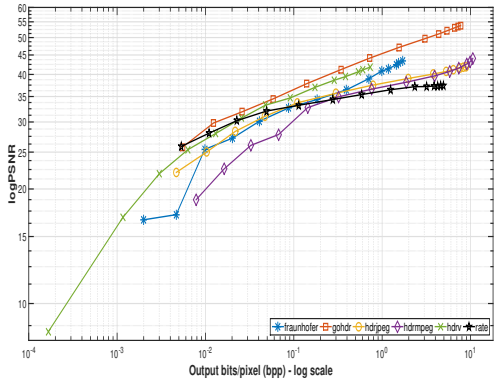


(g) HDR-VQM results (higher is better)

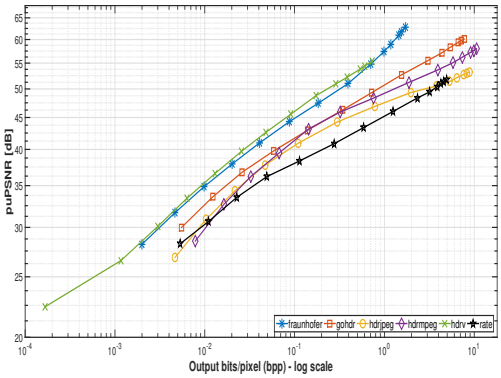
Figure 6.6: Averaged RD characteristics (quality vs output bitrate) of 39 HDR video compression algorithms against six QA metrics over 39 sequences.



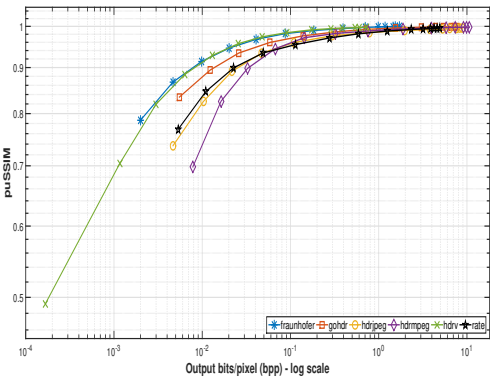
(a) PSNR results (higher is better)



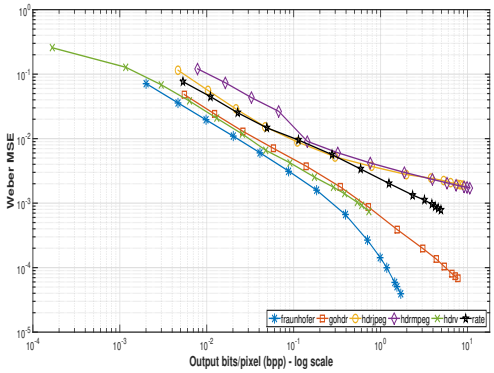
(b) logPSNR results



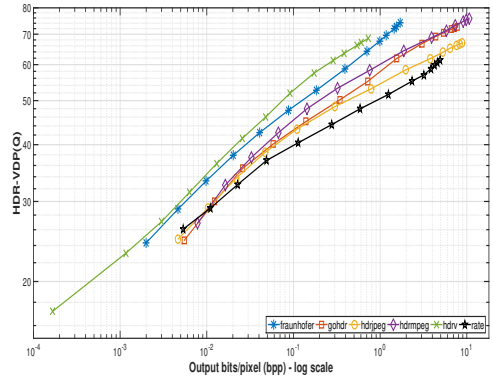
(c) puPSNR results



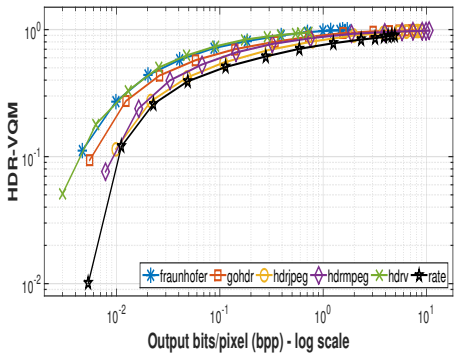
(d) puSSIM results



(e) Weber MSE results

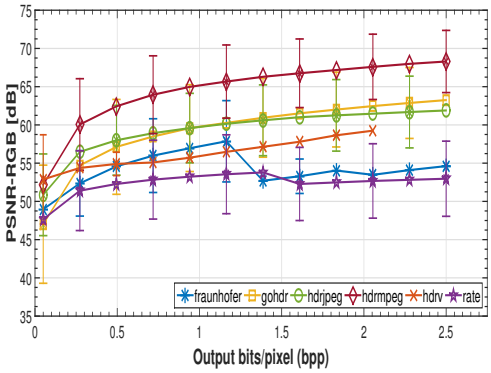


(f) HDR-VDP(Q) results

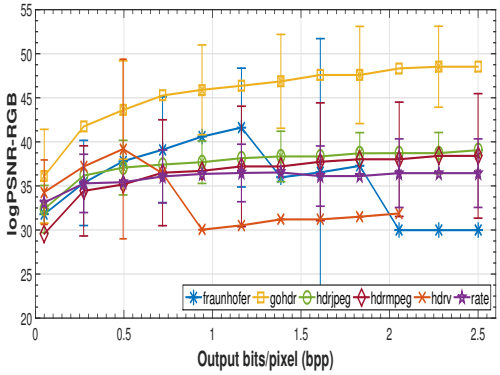


(g) HDR-VQM results (higher is better)

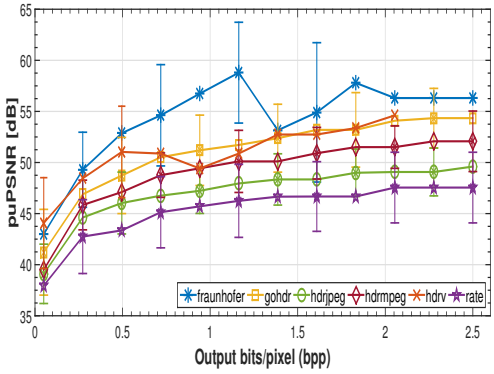
Figure 6.7: Averaged RD characteristics (quality vs output bitrate) of six HDR video compression algorithms against seven QA metrics over six short-listed sequences.



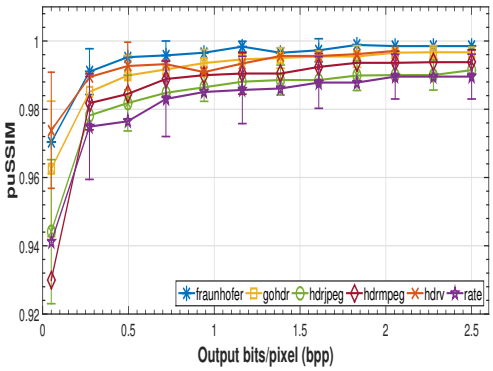
(a) PSNR results (higher is better)



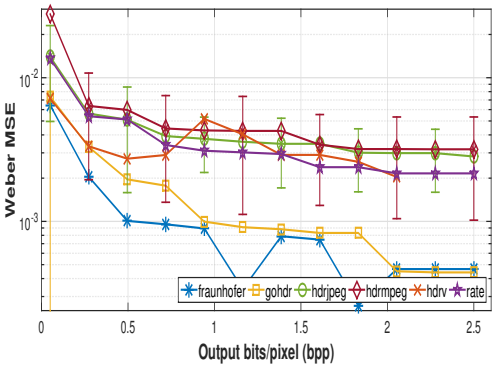
(b) logPSNR results (higher is better)



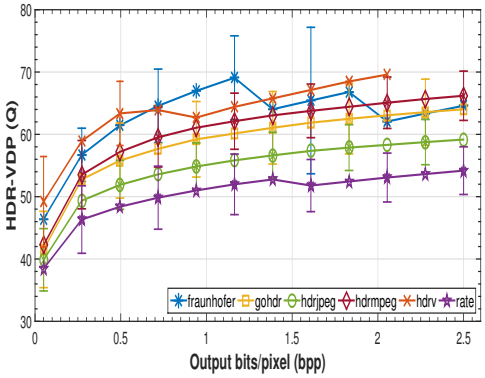
(c) puPSNR results (higher is better)



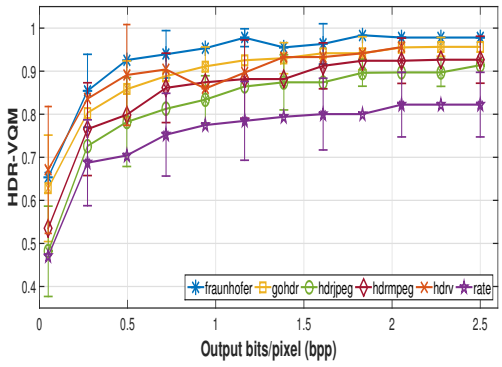
(d) puSSIM results



(e) Weber MSE results (lower is better)

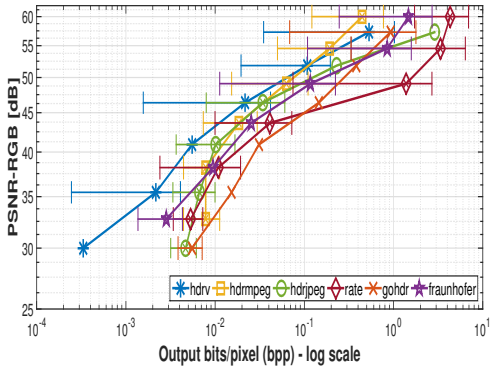


(f) HDR-VDP(Q) results (higher is better)

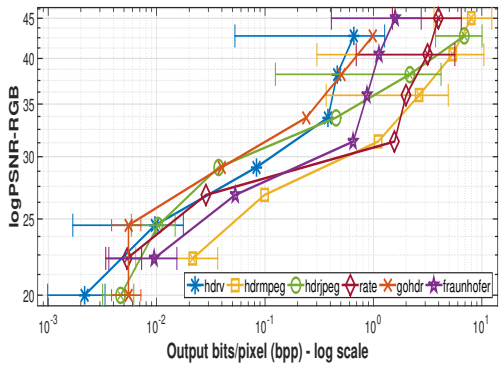


(g) HDR-VQM results (higher is better)

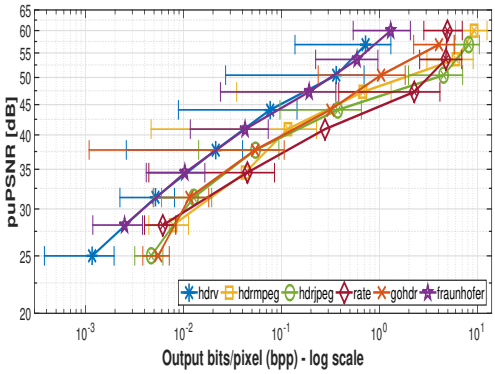
Figure 6.8: Interpolated RD characteristics for short listed sequences- fixed bitrates and interpolated quality levels.



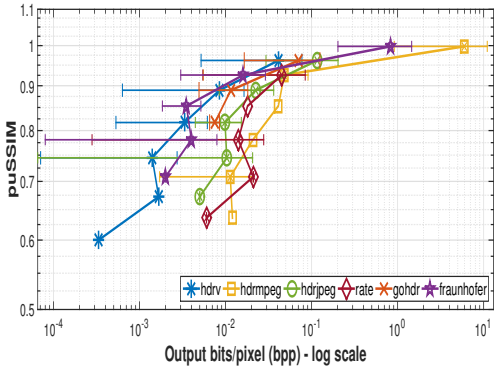
(a) PSNR results (higher is better)



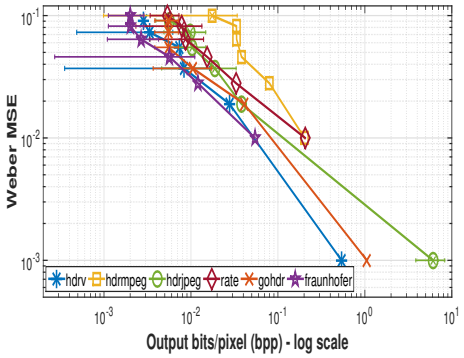
(b) logPSNR results (higher is better)



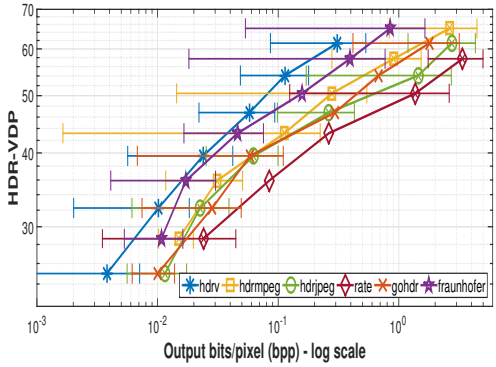
(c) puPSNR results (higher is better)



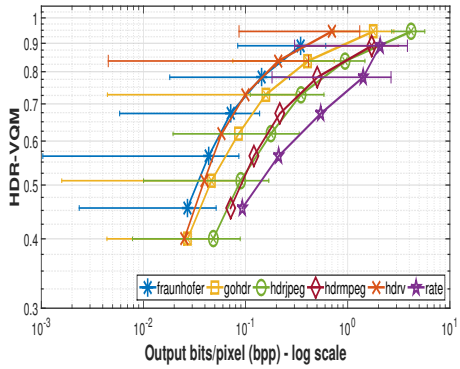
(d) puSSIM results (higher is better)



(e) Weber MSE results (lower is better)



(f) HDR-VDP(Q) results (higher is better)



(g) HDR-VQM results (higher is better)

Figure 6.9: Averaged RD characteristics (quality vs output bitrate) of six HDR video compression algorithms against seven QA metrics over six sequences. Results presented in log scale.

1. The RD characteristics of the algorithms exhibited against perceptual QA metrics such as *puPSNR*, *HDR-VDP* and *HDR-VQM*, as shown in Figures 6.4 and 6.7 demonstrate that *non-backward* compatible algorithms using higher bit-depth outperform their *backward* compatible counterparts at lower output bitrate.

2. Amongst the *non-backward* compatible algorithms, the *puPSNR, puSSIM, HDR-VDP* and *HDR-VQM* results exhibit that *hdrv* exhibits a superior HDR image reconstruction performance than *fraunhofer* and lower output bitrates. This is ratified both by Figures 6.5 and 6.8.

3. The inherent design of the *backward* compatible algorithms require a much higher bitrate to reproduce an acceptable image quality (without H.264 blocking artefacts). The mean output bitrate for *hdrjpeg* and *gohdr* (*backward* compatible algorithms - with residual streams containing the luma channel only) are similar to each other. The exceptions are *hdrmpeg* and *rate*. In *hdrmpeg*, both the base and residual streams contain 3-channels and are encoded with *High 4:4:4* sub-sampling. Again in *rate*, the Lagrangian optimization applied to the residual stream reduces the overall output bitrate, albeit at the cost of image quality.

6.5 Subjective evaluation

Most full reference QA metrics, were designed to evaluate image pairs without taking psychophysical aspects of the human visual system into consideration. Although perceptual QA metrics are good indicators of perceived image quality, the variation in objective results emphasizes the requirement for a comprehensive subjective evaluation.

6.5.1 Design

Multiple subjective evaluations at different image quality levels are ideally required to verify and correlate the results with objective evaluation. However, such an undertaking is very time consuming. Therefore, this work presents the results of two ranking-based psychophysical evaluations at two different quality levels. A ranking-based evaluation was chosen since it requires only one HDR display and guarantees that each ranked compression technique has a unique value, thereby ensuring quick and decisive results as opposed to a full-pairwise comparison experiment. Also, the relative rapidity of the process, approximately 20 minutes per participant, reduces fatigue.

The primary goal of the experiments was to rank and identify the order of each algorithm, across the six short-listed sequences, at two different quality levels. Participants were tasked to rank six algorithms for each of the six sequences, one at a time. For each sequence they had to view HDRVs from each algorithm at least once. They were tasked to identify and rank the given HDRVs in order of their resemblance to the clearly labelled

reference HDRV. Also a *hidden reference*, identical to the labelled reference was mixed with the algorithms.

The sequences and algorithms were randomly presented in order to avoid bias. While ranking the sequences, participants were allowed to view the HDRVs as many times as required. The motivation behind this was to be able to distinguish between HDRVs that are relatively close in quality without the exhaustive full-pairwise comparisons.

6.5.2 Materials

Software resources included HDRVs from six compression algorithms, uncompressed reference HDRVs and a graphical user interface (GUI) for the ranking-based experiment. Hardware resources included a SIM2 HDR display [SIMa] with a peak luminance rating of 4000 cd/m^2, an LG 22″ LED display with peak luminance rating of 300 cd/m^2 and a computer with a solid state drive for quick loading of HDRVs.

HDRVs for psychophysical experiment

Two fixed bpp(s) representing two quality levels were selected based on the objective results shown in Figures 6.7 and 6.8, respectively. The lower quality (LQ) level was chosen at 0.15 bpp ($\approx$ 8.8 Mbps - similar to online streaming quality), such that the image-quality distortions are clearly visible but not obscured by H.264 blocking artefacts. The higher quality (HQ) level was chosen at 0.75 bpp ($\approx$ 44.49 Mbps - similar to blu-ray quality).

Following the chosen quality levels, the six sequences were encoded at different QP settings for each algorithm to achieve the closest possible match to the target bitrate (within 5% error margin). Subsequently, the reconstructed HDR frames were converted to a custom file format suitable for displaying the HDR frames at 30 fps on a SIM2 HDR display. Table 6.2 demonstrates the target versus the achieved bitrate for each of the six algorithms along with the error margin.

Software for psychophysical experiment

A custom GUI application, shown in Figure 6.10, was specifically built for the ranking-based subjective evaluation. It presents seven thumbnails each linked to an HDRV (labelled A-G), six from different algorithms and a hidden reference for each sequence, on the left side of the screen. The clearly marked reference HDRV (or ground truth) thumbnail is presented in the centre. Each thumbnail, when *double-clicked* plays the linked HDRV on the HDR screen. Participants are tasked to view the reference HDRV first and subsequently rank the HDRVs on the left side in order of resemblance with the reference by dragging their preferred choice to its corresponding position (labelled 1-7) on the right side. The instructions for carrying out the experiment is clearly described in a text box below the reference thumbnail.

Algorithm	Target bpp	Achieved bpp	Error
hdrv	0.15	0.148	1.33%
hdrmpeg	0.15	0.159	6.00%
hdrjpeg	0.15	0.155	3.30%
rate	0.15	0.157	4.66%
gohdr	0.15	0.157	4.66%
fraunhofer	0.15	0.161	7.33%
Average	0.15	0.156	4.00%

(a) Target vs achieved bpp for the LQ experiment

Algorithm	Target bpp	Achieved bpp	Error
hdrv	0.75	0.71	5.30%
hdrmpeg	0.75	0.72	2.60%
hdrjpeg	0.75	0.76	1.33%
rate	0.75	0.77	2.66%
gohdr	0.75	0.76	1.33%
fraunhofer	0.75	0.76	1.33%
Average	0.75	0.74	1.33%

(b) Target vs achieved bpp for the HQ experiment

Table 6.2: Target vs achieved output bpp with error margin for lower and higher quality HDRVs

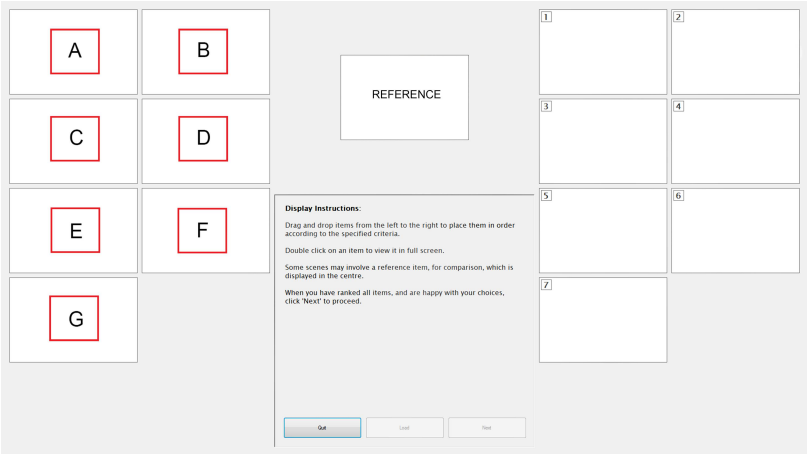


Figure 6.10: Screenshot of the evaluation software

6.5.3 Participants

A total of 64 participants were divided into two groups, 32 for each experiment (LQ and HQ), with an age range of 20 to 50 years and from various academic and corporate backgrounds took part in the experiments. The participants reported normal or corrected to normal vision.

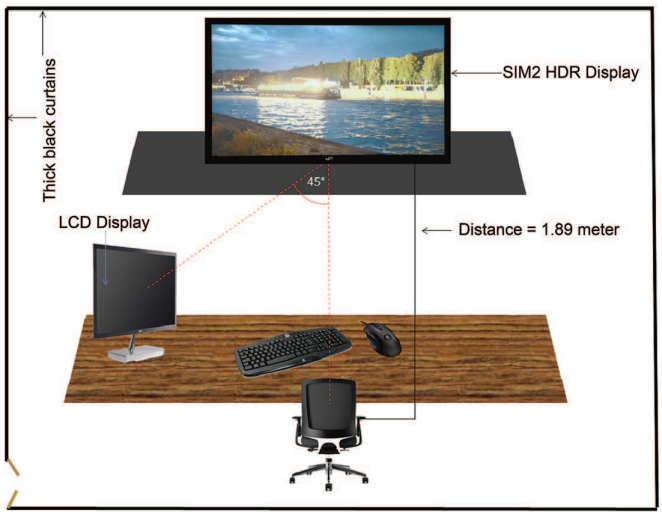


Figure 6.11: Psychophysical experiment setup.

6.5.4 Environment

Following ITU-R recommendations [ITU12], the experiments were conducted in a room with minimal ambient lighting (below 25 lux) which is within the recommended luminance levels for a typical dark environment [Eng]. The distance between the HDR display and the participant was set to approximately 3.2 times the height of the HDR display; at a distance of $\approx$ 189 cm with an LCD monitor placed at an angle of 45° (see Figure 6.11). In order to minimize glaring, the brightness and contrast of the LCD monitor was reduced to 25%.

6.5.5 Procedure

The participants were introduced to the objectives of the experiment prior to the start followed by a brief training session using a particular sequence subsequently discarded from the results. Upon completion of the training, the participants were asked to proceed further and rank the decoded HDRVs for the six sequences.

Each participant had to first view the reference HDRV on the HDR screen. Subsequently, the participant had to view each of the seven decoded HDRVs including the hidden reference and perform a qualitative assessment as to how much the decoded HDRVs resembled the ground truth HDRV in the centre. Based on their judgement, the participants positioned the corresponding thumbnails to one of the blank positions on the right, labelled [1-7], 1 being an HDRV with least distortion compared to the reference and 7, being the HDRV with most visible distortions.

6.5.6 Results

This section provides an overview of the results obtained from the psychophysical experiments and analyses the same.

Let the Null Hypothesis H_0 be that there are no significant differences between the compression algorithms for both LQ and HQ. The alternative H_1 states that there are significant differences between the algorithms. The statistical significance p is assumed to be 0.05. The sample size for both LQ and HQ is 32. Also, if H_1 is true, it is important to determine the coefficient of concordance which measures the degree by which the participants mutually agree on choices.

Now, let A(N, M, S) be a 3-dimensional data array where N denotes all participants, M denotes all compression methods (algorithms) and S denotes all six sequences. Therefore, A(N,M,S) represents the ranks given by each participant to each method for each of the six sequences.
This implies that $\bar{A}(\bullet,M,S)$ represents the mean ranks for each M and S, averaged across all participants. Also, $\bar{A}(\bullet,M,\bullet)$ represents the mean ranks averaged across all participants and sequences keeping M fixed. The grand average should be equal such that:

$$\frac{1}{K}\sum_{S=1}^{K}\bar{A}(\bullet,M,S) = \bar{A}(\bullet,M,\bullet) \text{ ,where K = total number of sequences.} \tag{6.4}$$

Furthermore, for each sequence S, assuming $N = 32$, being the number of participants and $M = 7$, being the number of methods (algorithms), let $r_{i,j}$ be the rank assigned to each algorithm $i \in M$ by each participant $j \in N$. Thus, for each algorithm i, $R_i = \sum_{j=1}^{N} r_{i,j}$ is the sum of the ranks assigned by N participants. $\bar{R} = \frac{1}{N}\sum_{j=1}^{N} r_{i,j}$ is the mean of the ranks assigned to algorithm i and $R = \sum_{i=1}^{M}(R_i - \bar{R})^2$ is the standard squared deviation. Finally, from this data, the Kendall's coefficient of concordance W can be computed as:

$$W = \frac{12R}{N^2(M^3 - M)} \tag{6.5}$$

The significance of W can be analysed using chi-squared statistics such that $\chi^2 = \frac{\mathrm{M(M-1)(1+W(N-1))}}{2}$. χ^2 is asymptotically distributed with $\frac{\mathrm{M(M-1)}}{2}$ degrees of freedom. A significance between scores suggests that the perceived image quality of two algorithms, when compared with each other are different although no conclusions can be drawn for cases of similarity.

The analysed results show that there are statistically significant differences between the algorithms for the six sequences. All tests show a significance $p < 0.05$. Therefore, H_0 is rejected and H_1 is accepted. Table 6.3a and 6.3b represents each $\bar{A}(\bullet,M,S)$ and $\bar{A}(\bullet,M,\bullet)$ along with W score and χ^2 value for the LQ and HQ experiments respectively. Furthermore, by combining the results of Tables 6.3a and 6.3b, a generalized $\bar{A}(\bullet,M,\bullet)$ can be computed for the entire subjective experiment across a total sample size $\mathrm{N_{total}} = 64$ as given in Table 6.3c.

Sequence	Compression Algorithms with Mean Rankings (@ 0.15 bpp)	Kendall (*W*)	X^2	*Sign. μ*
Welding	reference (1.53) hdrjpeg (2.63) fraunhofer (3.75) hdrv (3.78) gohdr (4.25) rate (5.97) hdrmpeg (6.09)	0.586	112.58	p < 0.01
CGRoom	reference (1.56) hdrv (2.13) fraunhofer (3.75) gohdr (4.47) hdrjpeg (5.28) hdrmpeg (5.41) rate (5.41)	0.548	105.16	p < 0.01
Jaguar	reference (1.37) hdrv (2.75) fraunhofer (3.68) hdrjpeg (3.71) hdrmpeg (4.59 gohdr (5.43) rate (6.43)	0.607	116.50	p < 0.01
Seine	reference (1.68) fraunhofer (2.81) hdrv (3.71) gohdr (3.78) hdrjpeg (4.34) hdrmpeg (5.18) rate (6.46)	0.518	99.48	p < 0.01
TOS	reference (1.75) hdrv (2.56) hdrmpeg (3.34) fraunhofer (3.46) hdrjpeg (5.00) rate (5.71) gohdr (6.15)	0.587	112.76	p < 0.01
Mercedes	reference (1.21) hdrv (2.93) fraunhofer (3.46) gohdr (3.84) hdrmpeg (4.56) hdrjpeg (5.15) rate (6.81)	0.669	128.46	p < 0.01
Ā(. , M , .)	reference (1.52) hdrv (2.97) fraunhofer (3.48) hdrjpeg (4.35) gohdr (4.65) hdrmpeg (4.86) rate (6.135)	0.783	150.26	P < 0.01

(a) Subjective ranks with Kendall *W*, averaged across participants at 0.15 bpp

Sequence	Compression Algorithms with Mean Rankings (@ 0.75 bpp)	Kendall (*W*)	X^2	*Sign. μ*
Welding	reference (2.75) fraunhofer (3.06) gohdr (3.12) hdrv (3.28) hdrmpeg (5.00) hdrjpeg (5.34) rate (5.43)	0.307	58.94	p < 0.01
CGRoom	reference (2.75) fraunhofer (3.00) hdrv (3.06) gohdr (3.71) hdrmpeg (4.93) rate (5.00) hdrjpeg (5.53)	0.277	53.10	p < 0.01
Jaguar	reference (2.28) fraunhofer (2.72) hdrv (3.00) gohdr (3.53) hdrmpeg (5.28) rate (5.46) hdrjpeg (5.71)	0.449	86.18	p < 0.01
Seine	hdrv (2.63) reference (2.81) fraunhofer (3.06) gohdr (3.40) hdrmpeg (4.93) hdrjpeg (4.93) rate (6.21)	0.400	76.88	p < 0.01
TOS	reference (2.91) hdrv (3.41) fraunhofer (3.46) hdrmpeg (3.62) gohdr (4.31) hdrjpeg (5.00) rate (5.28)	0.168	32.30	p < 0.01
Mercedes	reference (2.81) hdrv (3.46) gohdr (3.62) hdrmpeg (3.78) fraunhofer (3.93) rate (5.15) hdrjpeg (5.21)	0.168	32.27	p < 0.01
Ā(. , M , .)	reference (2.71) hdrv (3.14) fraunhofer (3.21) gohdr (3.61) hdrmpeg (4.59) hdrjpeg (5.29) rate (5.42)	0.511	98.20	P < 0.01

(b) averaged across participants at 0.75 bpp

Sequence	Compression Algorithms with Mean Rankings (combined for 0.15 & 0.75 bpp)	Kendall (*W*)	X^2	*Sign. μ*
Ā(. , M , .)	reference (1.66) hdrv (2.62) fraunhofer (3.02) gohdr (4.16) hdrjpeg (4.90) hdrmpeg (5.22) rate (6.42)	0.597	229.15	P < 0.01

(c) Subjective mean ranks with Kendall *W*, combined and averaged over HQ and LQ experiments

Table 6.3: Subjective results and groups for the LQ and HQ experiments

6.5.7 Analysis

Tables 6.3a and 6.3b illustrate the ranking of all methods. For those compression algorithms that are grouped together, no significant difference was found at $p < 0.05$. However, there are significant differences in-between separate groups. Larger groups, as seen mostly in Table 6.3b indicate ambivalence of participants in choosing one algorithm over another. The corresponding low Kendall W reaffirms the difficulty in comparing algorithms at higher output bitrates. However, large groups are less likely for the LQ experiment and higher Kendall W confirms the consistency in participants' choices; it is expected that more differences are noted at lower qualities. The combined results obtained from Table 6.3c further reduces the sizes of the groups providing a generalized subjective result with a moderately high degree of consistency $W = 0.597$ between the overall participants' choices.

Based on the mean ranks of each M from the three $\bar{A}(\bullet, M, \bullet)$ in Tables 6.3a, 6.3b and 6.3c, the algorithms can be assigned an ordinal rank. Such an ordinal ranking system as given in Table 6.4 presents a summarized information about the choices made by participants in the LQ and HQ experiments. Table 6.4 also presents the ordinal ranks when the LQ and HQ results are combined.

Algorithm	LQ ranking	HQ ranking	LQ + HQ
hdrv	1	1	1
fraunhofer	2	2	2
gohdr	4	3	3
hdrjpeg	3	5	4
hdrmpeg	5	4	5
rate	6	6	6

Table 6.4: Ordinal ranks for both LQ and HQ subjective experiments

6.6 Discussion

This discussion combines the objective and subjective results in order to establish a correlation between them and analyse the overall performance of the algorithms.

The reconstructed HDR frames at 0.15 bpp and 0.75 bpp from each algorithm are evaluated against the reference sequences using the previously mentioned QA metrics. The overall results from the algorithms for the six QA metrics can be sorted using the same ordinal ranking system as discussed in the previous section. Finally, a correlation is computed by combining the objective and subjective ordinal rankings at 0.15 bpp and 0.75 bpp using statistical non-parametric tests such as Spearman's rho rank correlation test. Table 6.5 shows the results from Spearman's rho rank correlation results for the combined LQ and HQ experiments.

First of all Table 6.5 shows that there are significant correlations between objective and subjective evaluation. For both the LQ and HQ experiments, the correlation between QA metrics, such as puPSNR/puSSIM/HDR-VDP/HDR-VQM and subjective rank-

	PSNR	logPSNR	puPSNR	puSSIM	Weber MSE	HDR-VDP	HDR-VQM	LQ	HQ
PSNR	-	-0.257	0.371	0.257	-0.232	0.493	0.257	0.371	0.371
logPSNR	-.257	-	.657	.771	.812*	.522	.771	.600	.657
puPSNR	.371	.657	-	.943**	.725	.986**	.943**	.829*	1.000**
puSSIM	.257	.771	.943**	-	.841*	.899*	1.00**	.943**	.943**
Weber MSE	-.232	.812*	.725	.841*	-	.632	.841*	.754	.725
HDR-VDP	.493	.522	.986**	.899*	.632	-	.812*	.899*	.986**
HDR-VQM	.257	.771	.943**	1.00**	.841*	.899*	-	.943**	.943**
LQ	.371	.600	.829*	.943**	.754	.812*	.943**	-	.829*
HQ	.371	.657	1.000**	.943**	.725	.986**	.943**	.829*	-

Table 6.5: Spearman's Rho rank correlation between objective and subjective evaluation for the LQ and HQ experiment respectively. '*' denotes significance at $p < 0.05$ level and '**' denotes significance at $p < 0.001$ level

ings is very high with statistical significance at $p < 0.001$ level. However, the correlation in-between the QA metrics varies with the image quality. While Table 6.5 shows very high correlation in-between puPSNR, puSSIM, HDR-VDP and HDR-VQM, the correlation in-between perceptual and mathematical metrics such as HDR-VDP and PSNR respectively, is significantly low. Finally, analogous to previous studies mentioned in Sections 4.2 and 4.5, PSNR demonstrates a significantly low correlation with subjective rankings.

The objective results suggest, that dedicated *non-backward* compatible algorithms tend to outperform their *backward* compatible counterparts at low to moderately high output bitrates. However, the differences are less clear to human participants. The $\bar{A}(\bullet, M, \bullet)$ groups in Table 6.3a include *fraunhofer* and *hdrjpeg* in a single group which suggests no significant difference between the algorithms at 0.15 bpp. Similarly, the $\bar{A}(\bullet, M, \bullet)$ groups in Table 6.3b include *gohdr* along with the *non-backward* compatible algorithms. Finally, the combined data in Table 6.3c shows that although *fraunhofer* and *gohdr* are part of the same sub-group, *hdrv* and *fraunhofer* are preferred over other *backward* compatible algorithms.

It is important to note that even though *non-backward* compatible algorithms perform well at lower bpp, the output streams cannot be played back using existing video players. Furthermore, in practice, hardware support for 10/12 bit encoders and decoders are currently quite rare. However, the flexibility and simplicity of the *fraunhofer* design facilitates an easier adaptation to upcoming 10-bit video pipelines. On the other hand, the *backward* compatible algorithms can use existing 8-bit video pipelines providing a distinct advantage in early adoption of HDR.

Out of the four *backward* compatible algorithms, *hdjpeg, hdrmpeg* and *rate* contain an 8-bit tone-mapped base stream making them truly backward compatible. *gohdr*, the only exception, is able to match the performance of *non-backward* compatible algorithms, albeit at the cost of true backward compatibility. Also, the Lagrangian optimization in *rate* saves output bitrate at the cost of reconstructed image quality. Although, Lee et al. [LK08] claimed, that *rate* performed better than *hdrmpeg* at lower bitrates, the algorithm was tested on sequences with VGA resolution and an older version of HDR-VDP [MMS04]. There-

fore, the claim might not always hold true for a large set of full HD resolution sequences.

The overall comparison results suggest that the choice of compression algorithm is largely application specific. The best possible HDR video quality at minimal output bitrates can be delivered by *non-backward* compatible algorithms. However, video pipelines with higher bit-depth support are required in that case. On the other hand, *backward* compatible algorithms can deliver HDR video content using legacy pipelines albeit at the cost of higher output bitrates. The choices are reaffirmed by the algorithms' RD characteristics against perceptual QA metrics which evidently has a high correlation with the subjective evaluation.

6.7 Conclusion

This work endeavours to provide a detailed comparison of a number of published and patented HDR video compression algorithms and forms the foundation against which other HDR video compression algorithms can be evaluated in future. It establishes that *non-backward* compatible compression algorithms enjoy a distinct advantage over their *backward* compatible counterparts at lower bitrates. Also, *backward* compatible algorithms require *dual-loop* encoding to create the residual stream. This adds more complexity to any hardware design thereby making it less suitable for real-time deployment.

The work presented in this chapter opens up avenues of future research. In practice, 12-bit hardware encoders and decoders are not available to date and this work presents a comprehensive evaluation of compression algorithms which were proposed before the MPEG CfE for HDR/WCG compatibility with HEVC-Main-10 profile. Therefore, an interesting research area which has recently gained traction following the HDR/WCG call for proposals to would be to see how *non-backward* compatible compression algorithms (including the ones presented in this work) can be adapted to perform with HEVC Main-10 profile [SOHW12] for even lower bitrates. Subsequently, the modified non-backward compatible algorithms can be evaluated against the recently adopted Perceptual Quantizer algorithm (SMPTE ST 2084) [MND13] and Hybrid Log-Gamma algorithm [BC15] to test and compare their HDR reconstruction performance such as has been considered by François et al. [FFH*16]. Efficient use of available bandwidth might finally lead to the widespread commercial adoption of HDR.

6.8 Summary of the design decisions

This chapter describes a comprehensive objective and subjective evaluation of some of the most relevant HDR video compression algorithms and in doing so establishes a methodology to evaluate and understand the advantages and disadvantages of the two different approaches to HDR video compression as well as of the individual algorithms. The summary of the design decisions and parameters which can be inferred from the best performing

algorithms (according to the combined ordinal ranking) are as follows:

- The *non-backward* compatible approach provides better reconstruction quality at significantly lower transmission cost compared to the *backward* compatible algorithms. Therefore, the *non-backward* compatible approach is likely to be the preferred approach unless constrained by hardware and software requirements (higher bit-depth requirments).

- Separation of luminance and chroma information required for effective manipulation of the luminance information. Since the HVS is less susceptible to chroma information loss, the preservation of luminance information is prioritised over the preservation of chroma information. Also, usage of uniform chromaticity scale based colour spaces such $Yu'v'$ is preferred over the traditional YC_bC_r colour space.

- PTFs based on contrast sensitivity functions (CSFs) can be used to encode luminance information to JND spaced luma code values. Additionally, subjective evaluation based PTFs such as Ferwarda's *t.v.i* used in *hdrv* provides a more efficient encoding of luminance information compared to the logarithmic PTF used in *fraunhofer*.

- Usage of metadata information is critical as used by both *non-backward* compatible and *backward* compatible HDR video compression algorithms. The metadata information is required for accurate reconstruction of HDR video frames.

- Although, *gohdr* performs marginally better other *backward* compatible algorithms, it is preferred to design a *backward* compatible algorithm where the base stream is a tone-mapped representation of the HDR video sequence since this allows the stream to be played using legacy video players.

Based on the knowledge and understanding gained from the derived results, the next chapter in this thesis introduces a novel *non-backward* compatible compression algorithm which delivers better performance that the existing state-of-the-art algorithms.

Chapter 7

Uniform Colour Space based HDR Video Compression

CHAPTER 6 provided a comprehensive objective and subjective evaluation of existing *non-backward* and *backward* compatible HDR video compression algorithms. This facilitates a comprehensive understanding of the design decisions behind each approach and each compression algorithm. Additionally a detailed benchmarking methodology led to a short-listing of the best HDR video compression algorithms and facilitates a detailed understanding of the advantages and shortcomings of the existing state-of-the-art. It also established that *non-backward* compatible HDR video compression algorithms deliver better image reconstruction quality at lower transmission cost compared to *backward* compatible solutions.

To satisfactorily answer our research question, this Chapter proposes a novel HDR video compression algorithm which endeavours to provide better image reconstruction quality at lower transmission and storage costs than existing solutions. The proposed algorithm is the third and final step of answering the primary research question. The algorithm described in this Chapter uses a hitherto unused perceptually uniform colour opponent *Intensity, Protan and Tritan* (IPT) space [EF98a], proposes a novel PTF to encode the dynamic range of the scene and introduces a new Error Minimisation Function (EMF) for accurate chroma reproduction. In addition to the proposed hybrid PTF and chroma EMF, the proposed algorithm allows the use of any existing PTFs/OETFs to encode the scene dynamic range to JND scaled luma values. This provides a degree of flexibility hitherto unavailable in HDR video compression. An overview of the proposed algorithm is given in Figure 7.1.

The proposed algorithm has been evaluated against four state-of-the-art published and/or patented *non-backward* compatible HDR video compression algorithms using a set of 39 HDR video sequences, the latest x265 [Orgb] (an HEVC [SOHW12] implementation) codec at 11 different quality levels against the same seven full-reference objective QA metrics (mentioned earlier in Chapter 6). The evaluation data provides a set of generalised

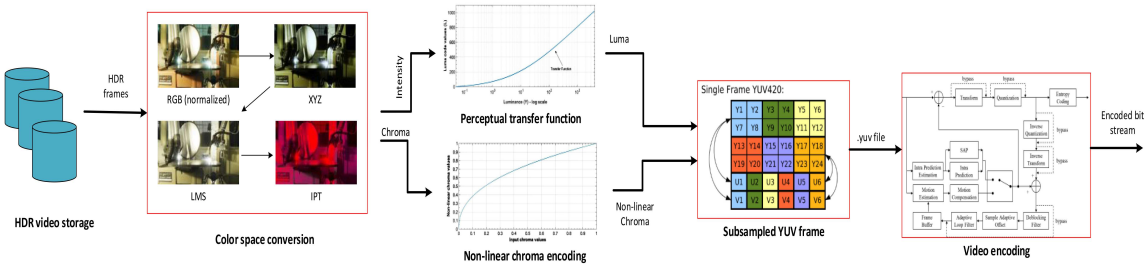


Figure 7.1: An overall workflow of the proposed HDR video compression algorithm.

RD characteristics which is used for overall benchmarking of the five algorithms (including the proposed). The raw data obtained from the objective evaluation was subsequently interpolated for an in-depth study of the compression performance. Results demonstrate that the proposed algorithm exhibits the least coding error amongst the five algorithms evaluated. Additionally, RD characteristics suggest that the proposed algorithm outperforms the existing state-of-the-art at bitrates ≥ 0.4 bits/pixel.

The primary contributions of this work are:

1. A novel *non-backward* compatible HDR video compression algorithm which uses a combination of IPT color opponent space, a novel PTF to encode scene dynamic range and a new EMF to non-linearly encode chroma information.

2. A modular structure of the algorithm to use existing any contrast sensitivity function (CSF) based PTF inside the algorithm's *intensity* encoding block to encode the scene dynamic range to JND quantised luma space.

3. A comprehensive objective evaluation of the proposed algorithm against four existing state-of-the-art algorithms for performance benchmarking purposes.

7.1 Background

This section provides an overview of some of the underlying concepts based on which the proposed algorithm has been designed.

7.1.1 Colour spaces

HDR data is generally stored in linear RGB format which is highly device dependent and has a high correlation in-between the channels [RKAJ08]. To minimise the effect of pixel manipulation on one channel affecting the others, RGB pixel values are typically converted to luma-chroma spaces such as YC_bC_r or $Yu'v'$ where u' and v' represent uniform chromaticity scales. Also, for efficient compression purposes perceptual uniformity is desirable where the perceived difference in-between two colors is equal to the Euclidean distance between them [RKAJ08]. Although the CIE-XYZ space can be used, it is not perceptually uniform and contains imaginary primaries with a large number of values which do not correspond to realisable colors leading to an inefficient use of available bit-depth [RKAJ08]. Therefore,

existing algorithms [MKMS04, GT11, MND13, BC15] have used luma-chroma spaces such as the YCbCr the extended Yu′v′ space. However, these color spaces are again not perfectly uniform. Thus, to address both the essential and desirable properties, the RGB data can be converted to device independent *hue, saturation and lightness* (HSL) color opponent spaces such as CIELAB/LUV [Fai13a]. However, further research [HB95, EF98b] have confirmed issues with hue compressibility in CIELAB/LUV. Thus, the proposed algorithm uses the IPT color opponent space [EF98a] which maintains the perceptual uniformity of CIELAB/LUV and mitigates the hue compressibility issues. Further details about the usage is discussed later in Section 7.2.2.

7.1.2 Perceptual Transfer Functions

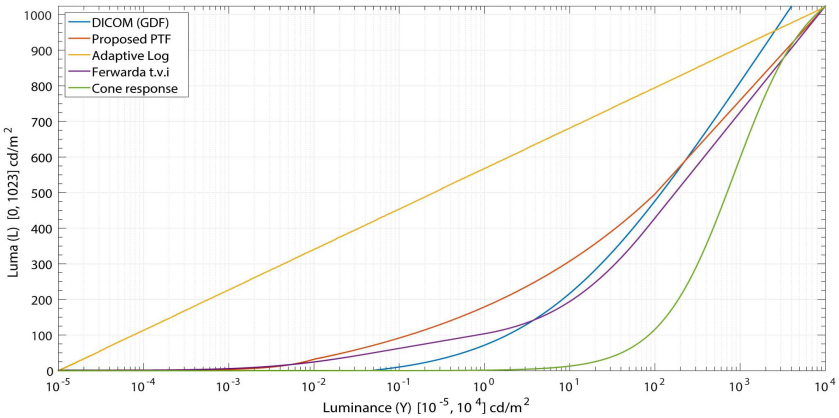


Figure 7.2: A log-linear plot of five perceptual transfer functions (including a novel proposed PTF).

In HDR video compression, a transfer function (TF) is ideally a reversible function which maps a range of input pixel values R_i to a range of output code values R_o such that R_o is suitable for video encoding. A detailed discussion, derivation and requirement of PTFs/OETFs in HDR video compression has been discussed earlier in Section 3.2. As mentioned previously, the proposed algorithm has been designed in the form a framework (see Section 7.2 for details) such that several existing PTFs can be used to encode the dynamic range to JND scaled luma values.

This section provides a brief overview and discusses the characteristics of four established PTFs which have been used in conjunction with the proposed algorithm to non-linearly encode the scaled *intensity* channel values as described later in section 7.2.3. Furthermore, the image reconstruction quality of the PTFs were compared with each other to determine the advantages,disadvantages and suitability of each PTF for HDR video compression purposes. The considered PTFs were the Adaptive LogLuv TF [MT10], the DICOM standard Grayscale Display Function [MDG08], the *t.v.i* proposed by Ferwarda et al. [FPSG96] with Ward's modification [War08] and the cone response model [NBRA83]. Based on an in-depth understanding of the established PTFs and comparative results, a

novel hybrid PTF was proposed to non-linearly encode the scaled intensity information, the details of which is outlined later in Section 7.2. The proposed hybrid PTF was subsequently compared with the established PTFs and the comparative results along with subsequent discussion is given in Sections 7.4 and 7.5, respectively.

Adaptive Logarithmic TF

A modification of a logarithmic function proposed by Motra and Thoma [MT10] in the form of an Adaptive LogLuv TF which adjusts and scales the output values based on the input and output boundary conditions. This enables the TF to encode the entire range of visible luminance into n − bit code values as shown in equation 7.1.

$$L = \left\lfloor \frac{2^n - 1}{log_2\left(\frac{Y_{max}}{Y_{min}}\right)} (log_2(Y) - log_2(Y_{min})) \right\rfloor \tag{7.1}$$

A logarithmic PTF, as shown in Figure 7.2 exhibits conservative quantisaton of lower luminance pixel values and coarser quantisation at higher luminance values. This can be attributed to the shape of the curve where a steeper curve results in a finer quantisation [MMS06]. However, previous psychophysical experiments have shown that the contrast detection thresholds of the HVS at scotopic and mesopic ranges are higher than at photopic luminance ranges. Therefore, the use of a logarithmic TF (also seen in LogLuv encoding [Lar98]) results in an inefficient usage of available bit depth [MMS06].

Grayscale Display Function

The DICOM standard Grayscale Display Function (GDF) [MDG08], a polynomial fit, derived from Barten's CSF experiments [Bar92] maps the input luminance $Y \in [0.05, 4000]$ cd/m^2 to a 10-bit perceptually uniform JND space using equation 7.2.

$$\begin{aligned} L = {} & A + B \cdot log_{10}(Y) + C \cdot (log_{10}(Y))^2 + \\ & D \cdot (log_{10}(Y))^3 + E \cdot (log_{10}(Y))^4 + F \cdot (log_{10}(Y))^5 + \\ & G \cdot (log_{10}(Y))^6 + H \cdot (log_{10}(y))^7 + I \cdot (log_{10}(Y))^8 \end{aligned} \tag{7.2}$$

Although the GDF is suitable for existing high-fidelity commercial displays, it is limited to 4000 cd/m^2 and future displays might exceed the encoding capabilities of this function. Also, the GDF exhibits exceedingly coarse quantisation below 1000 cd/m^2 and redundantly conservative quantisation for higher luminance values which renders it unsuitable for accurate scotopic and mesopic luminance preservation.

Ferwarda's t.v.i

Ferwarda et al. [FPSG96] proposed another *t.v.i* function which takes into account the non-linear response of rods and cones separately. The proposed TVI function models input luminance $Y \in [10^{-6}, 10^{9})$ cd/m^2 to a JND space for rods and cones separately. Although,

the HVS response as characterised by this *t.v.i* can be fitted using a double-exponential function, the responses can be approximated (by curve fitting) to create a single TVI function as shown in equation 7.3.

$$L = \begin{cases} -2.86 & \text{if } log_{10}(Y) \leq -3.94 \\ (0.405 \cdot log_{10}(Y) + 1.6) & \\ \times 2.18 - 2.86 & \text{if } log_{10}(Y) \in [-3.94, -1.44) \\ log_{10}(Y) - 0.395 & \text{if } log_{10}(Y) \in [-1.44, -0.0184) \\ (0.249 \cdot log_{10}(Y) + 0.65) & \\ \times 2.7 - 0.72 & \text{if } log_{10}(Y) \in [0.0184, 1.9) \\ log_{10}(Y) - 1.255 & \text{if } log_{10}(Y) > 1.9 \end{cases} \tag{7.3}$$

However, the *t.v.i* function is based on data from only 18 subjects and the detection thresholds are higher for low luminance values and banding artefacts might be visible due to the fact that the authors used a pulsating target on a constant background and perception thresholds are higher for transient stimuli compared to static stimuli [War08]. Therefore, Ward proposed a modification where the threshold luminances are divided by a factor of nine by subtracting 0.95 from the formula given in equation 7.3.

$$D(L) = 10^{tvi(L) - 0.95} \tag{7.4}$$

where $D(L)$ denotes the modified detection thresholds and $tvi(L)$ represents equation 7.3. According to Ward [War08], dividing the threshold by a factor of nine brings the function in better agreement with the Barten model and yet preserves detail below 10^{-2}cd/m^2.

Global Cone Response Model

The final PTF in consideration was the Global Cone Response Model (GCRM) [NBRA83] primarily targeted to model the HVS response at photopic levels. The model assumes that all cones of the HVS are adapted to the same luminance level and can be approximately formulated using equation 7.5

$$L = \frac{c_1 \cdot Y}{Y + (17.4) Y_{mean}^{0.63}} + c_2 \tag{7.5}$$

where:

$$\begin{aligned} Y_{mean} &= \overline{Y} \\ c_1 &= \frac{L_{max} - L_{min}}{\frac{Y_{max}}{Y_{max} + 17.4 \cdot Y_{mean}^{0.63}}} - \frac{Y_{min}}{Y_{min} + 17.4 \cdot Y_{mean}^{0.63}} \\ c_2 &= Y_{min} - \frac{c_1 \cdot Y_{min}}{Y_{min} + 17.4 \cdot Y_{mean}^{0.63}} \end{aligned}$$

where $Y \in [10^{-5}, 10^{4}]$ and $L \in [0, 2^{n} - 1]$.

The shape of GCRM (see Figure 7.2) indicates a conservative preservation of high luminance values at the cost of lower luminance values. In addition to the mentioned functions, several other PTFs have also been proposed to date. A detailed overview and derivation of several PTFs along with their effect on the visibility of contouring artefacts are given in [SYD87, MMS06], respectively.

7.2 Overview of the proposed algorithm

In this section, we provide a detailed overview of the proposed algorithm and highlight its several design aspects. The major contributions of the proposed algorithm are

1. The usage of the IPT color opponent space.
2. The proposal of a novel PTF (optimised for 10-bit encoding) with a straightforward analytical solution to perceptually encode the *intensity (I)* channel information.
3. The proposal of a novel error minimisation function (EMF) (optimised for 10-bit encoding) to accurately preserve the chroma information.

7.2.1 Overall data-flow

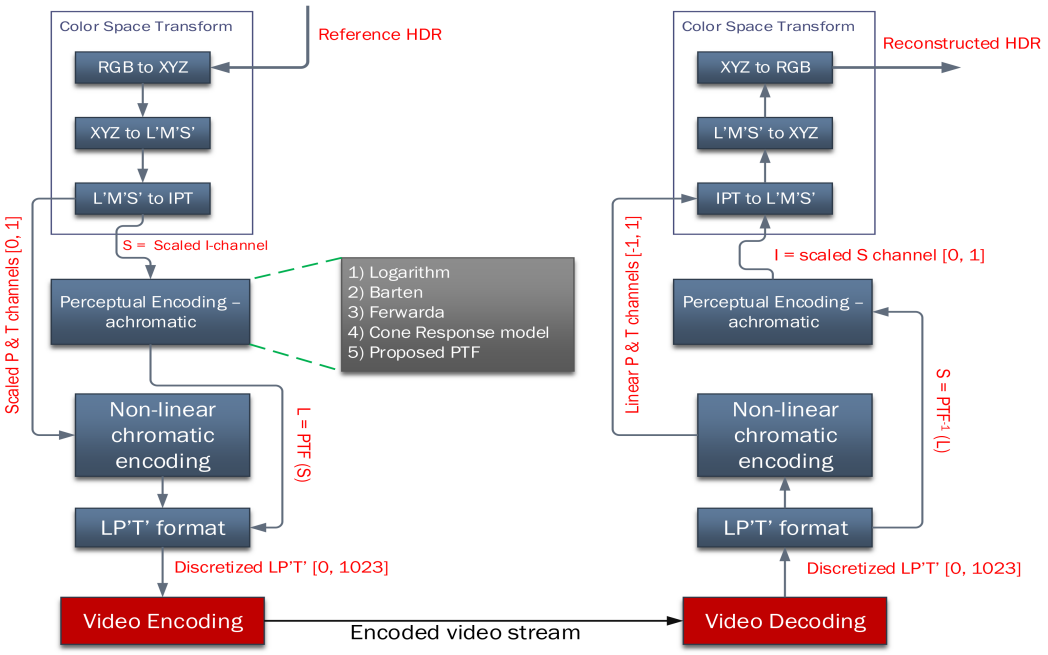


Figure 7.3: Schematic diagram of the proposed algorithm and framework

The proposed algorithm can be broadly classified into three modules. The first module normalises and performs colour space transform of input HDR frames (linear RGB) to perceptually uniform IPT colour opponent space (see Section 7.2.2 for details). The

second module extracts the *intensity* channel information from the resultant IPT frame and linearly scales the *intensity* information according to the requirements of a chosen PTF. The scaled *intensity* values are then perceptually encoded to JND scaled luma code values using the chosen PTF (see Section 7.2.3 for details). The third module extracts the chroma components from IPT and applies the EMF (see Section 7.2.4 for details) to non-linearly encode the chroma components. Finally, the luma and chroma components are merged and passed to the video codec for encoding.

On the decompression side, the encoded video stream is decoded and decompressed to reconstruct the HDR frames by reversing the data flow. The algorithm also uses metadata information (see Section 7.2.5 for details) which is used to accurately reconstruct the HDR frames. A visual description of the overall data-flow is given in Figure 7.3.

7.2.2 Module 1: Colour space transform

The psychophysical data available from [HB95, EF98b] demonstrated that widely used perceptually uniform colour spaces such as CIELAB and CIELUV cannot fully de-correlate the light and color information required for effective manipulation of HDR frames. Also, both CIELAB/LUV suffer from compressibility issues such as the hue changes that occur when compressing chroma along the lines of hue [RKAJ08].

To mitigate the limitations of CIELAB/LUV, the proposed algorithm first normalizes the input HDR frame (in linear RGB) and converts the normalized RGB to the IPT color opponent space [EF98a]. This transforms input data into a perceptually uniform space for ease of image manipulation and also de-correlates the light and color information for compression purposes. Therefore, the IPT color space enjoys all the advantages of CIELAB/LUV without the hue compressibility issues of CIELAB/LUV [RKAJ08]. A brief outline of the colour space transformation is given in Algorithm 1 and the details are given in Section 2.4.2.

Algorithm 1 ColourConvert(hdr)

1: $\nu \leftarrow max(hdr)$ //get normalisation factor
2: $rgb_{norm} \leftarrow \frac{hdr}{\nu}$ //normalisation
3: $IPT \leftarrow function(\text{rgb}_{\text{norm}}$ to IPT) //see Section 2.4.2 for details.
4: $P \leftarrow IPT(x,y,2)$ //extract the 2^{nd} channel
5: $T \leftarrow IPT(x,y,3)$ //extract the 3^{rd} channel
6: $P_{scale} \leftarrow \frac{P-min(P)}{max(P)-min(P)}$ s.t $P \in (0,1]$ //P scaling
7: $T_{scale} \leftarrow \frac{T-min(T)}{max(T)-min(T)}$ s.t $T \in (0,1]$ //T scaling
8: $IPT_{out} \leftarrow I(x,y,1), P_{scale}, T_{scale}$ //scaled IPT space

In Step 4, HPE refers to the Hunt-Pointer-Estevez fundamentals [NHTS87] and the metadata includes the normalisation factor ν along with the minimum and maximum pixel

values of the P and T channels prior to scaling.

7.2.3 Module 2: Perception based intensity encoding

Module 2 extracts the *intensity* channel from IPT_{out} (see Figure 7.3). The *intensity* information when linearly scaled and discretised for 10-bit encoding exhibits visible contouring artefacts due to rounding errors. Thus, to minimise the quantisation errors, the scaled I' channel was perceptually encoded by any of the previously mentioned PTFs such that the resultant JND quantised luma satisfies the properties mentioned in Section 7.1.2. Since $I \in (0,1]$, it can be scaled to any range, suitable for a chosen PTF. The linear scaling operation is performed by a multiplying factor ψ followed by the application of the PTF as shown in equation 7.6.

For instance, if $I \in (0,1]$, $f(\cdot)$ is the chosen PTF (say GDF) and L is the 10-bit JND quantised luma then the scaling and JND mapping operation is given as in equation 7.6.

$$\begin{aligned} I' &= I \cdot \psi \text{ such that } I' \in [0.05, 4000] \\ \therefore L &= f(I') \text{ such that } L \in [0, 1023] \end{aligned} \tag{7.6}$$

To determine the *intensity* encoding efficiency of each PTF and to evaluate the reconstruction quality of the algorithm (as a whole) upon the application of the PTF, the scaled I' channel is encoded using each of the four existing PTFs (one at a time). The rest of the data flow remains unchanged (see Figure 7.3). Subsequently, the algorithm is used to determine the reconstruction quality of the 39 HDR sequences using the evaluation methodology described later in Section 7.3.2. The RD characteristics across a set of different quality levels determines the overall HDR reconstruction quality of the algorithm when using each of the four PTFs. This indirectly indicates the *intensity* channel encoding efficiency of each PTF.

The RD characteristics discussed later in Section 5.5 show that amongst the existing PTFs, the algorithm exhibits the best reconstruction quality using either GCRM or the modified Ferwarda's *t.v.i.* However, both PTFs have certain issues as previously discussed in Section 7.1.2. Further details about the shape and characteristics of the PTFs have been discussed previously in Section 3.2.1. To mitigate those issues, this Chapter proposes a novel PTF which incorporates the advantages of both along with the added advantage of a straightforward analytical solution.

Design of the proposed PTF

Following recommendation REC 1886 [Ser11], the proposed PTF has been designed as a three-part analytical solution such that $f(\cdot) : I' \longrightarrow L$. The conditional equation 7.7 bears similarity to sRGB-non-linearity with linear and power function segments but additionally

includes a logarithmic segment to encode high *intensity* values.

$$L = \begin{cases} a \cdot I' & \text{if } I' < I'_s; \\ b \cdot I'^{(\frac{1}{c})} + d & \text{if } I' \in [I'_s, I'_p); \\ e \cdot log_{10}(I') + f_c & \text{if } I' \in [I'_p, I'_h]; \end{cases} \tag{7.7}$$

Similarly, $f^{-1}(\cdot)$ can be formulated as in equation 7.8.

$$I' = \begin{cases} \frac{L}{a} & \text{if } L < L_s; \\ (\frac{L-d}{b})^c & \text{if } L \in [L_s, L_p); \\ 10^{(\frac{L-f_c}{e})} & \text{if } L \in [L_p, L_h]; \end{cases} \tag{7.8}$$

The boundary value conditions I' was assumed to be similar to [MND13]. Therefore, $I \in (0,1]$ is scaled by ψ such that $I' \in [10^{-5}, 10^4]$. Also, the JND quantised $L \in [0, 1023]$. The goal of the proposed PTF was to facilitate a conservative quantisation throughout the range of I' for low-, mid- and high-*intensity* regions. Since the shape of GCRM shows biasedness towards preservation of high-*intensity* regions, it was taken out of consideration. Now, amongst the existing PTFs, the shape of Ferwarda's *t.v.i* is a very close fit to the analytical model proposed in Daly's VDP [Dal92] for the power segment and also a close fit to Barten's CSF based PTF for the logarithmic segment. Therefore, the proposed analytical model was initially fitted to Ferwarda's *t.v.i* using non-linear regression techniques for initial calculation of the interval boundaries I'_s and I'_p. I'_h was always fixed to 10^4 as the upper bound of the *intensity*, considered in this work. Such an excercise produces the intial interval boundaries as well as the co-factors in equation 7.7. Using the co-factors and interval boundaries, L_s and L_p were computed. Similar to I'_h, L_h was again fixed to 1023 as the upper bound for 10-bit encoding.

Since the analytical model is a piecewise-nonlinear model, it is important to enforce C^0 continuity at the intervals bounds I'_s and I'_p. Also, it is important to test the function for large jumps and discontinuities using a Contrast vs. Intensity (*c.v.i*) plot and corresponding adjust the parameters to not only enforce C^0 continuity but also eliminate jumps and discontinuities. This ensures the elimination of undesirable visible contouring artefacts. Therefore, following the completion of equation 7.7, the analytical model was re-verified using a *c.v.i* plot and discontinuities and large contrast jumps were found especially at the low-*intensity* regions. Correspondingly after iterative trials, the co-factors in equation 7.7 were adjusted to elimnate the contrast jumps and discontinuities. Using a *c.v.i* plot also provides an advantage in measuring the effectiveness of the bit-depth allocation in L. Re-plotting the *c.v.i* with the proposed PTF's modified co-factors showed that the bit-depth allocation was not optimal. Therefore, a second round of optimisation was performed on both the boundary values and co-factors to ensure optimal bit-depth allocation with re-modified

co-factors to eliminate contrast jumps and discontinuities. Furthermore, the C^0 continuity was re-enforced thus arriving to the final configuration of the proposed PTF in equation 7.7. Correspondingly the co-factors of equation 7.8 was computed. The interval boundaries and co-factors are given in Table 7.1.

$a = 2285.712$	$b = 224.1745$	$c = 5$
$d = -67.1009$	$e = 263.5$	$f_c = -31$
$I'_s = 0.007$	$I'_p = 100$	$I'_h = 10^4$
$L_s = 16$	$L_p = 496$	$L_h = 1023$

Table 7.1: Co-factors used for the proposed PTF.

This PTF when plotted with the final configuration, shows interval boundaries I'_s, I'_p and I'_h in equation 7.7 represent the brightness (in this case the scaled *intensity* channel) values where the HVS exhibits linear, power and logarithmic response, respectively [SYD87]. Correspondingly, the *c.v.i* plot ensured that the JND space L was divided into three blocks with optimal bit-depth allocation within intervals where $L \in (0, L_s)$, $L \in [L_s, L_p)$ and $L \in [L_p, L_h]$ such that each block can facilitate a conservative quantisation of low-, mid- and high-*intensity* regions. Also, when the modified PTF was plotted with a semilog plot (I' vs. L) and compared with the existing PTFs, the shape of the curve showed the following characteristics:

- In the low-*intensity* regions, the curve exhibits an optimal quantisation where it performs more conservative quantisation than exhibited by the modified Ferwarda's *t.v.i* while not as conservative as a logarithmic PTF.
- In the mid-*intensity* regions, the curve exhibits a similar quantisation to the modified Ferwarda's *t.v.i*.
- In the high-*intensity* regions, the curve exhibits a conservative quantisation similar to Barten's CSF based PTF.

Furthermore, the bit-depth allocation effectiveness was tested against existing EOTFs and found to be a close fit with the EOTF used in the PQ algorithm [MND13]. The *c.v.i* plot is given Figure 7.4. For a further confirmation, the proposed PTF and its inverse were rigorously tested by using them in the algorithm and evaluating the same using the same methodology described in Section 7.3. Results obtained from the evaluation showed that the performance of the proposed algorithm using the proposed PTF is better than the existing PTF used for the *intensity* channel encoding. The evaluation results are given later in Figure 7.7.

7.2.4 Module 3: Error minimisation function (EMF)

Similar to Module 2, this module extracts the chroma information (P & T channels) from IPT_{out} and performs a non-linear encoding which minimises quantisation errors frequently

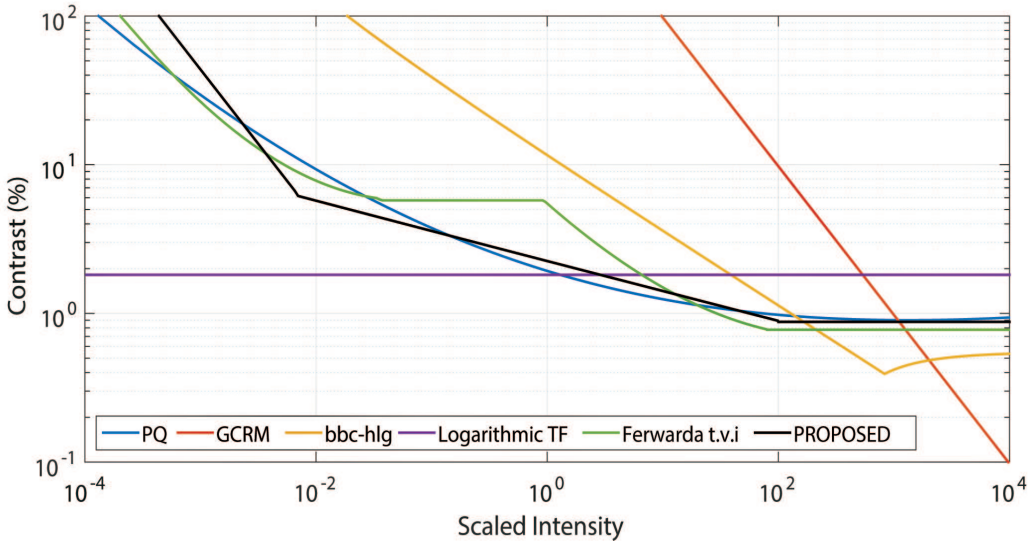


Figure 7.4: Comparative Contrast vs. Intensity plot of the proposed PTF compared to existing PTFs and EOTFs used in other algorithms.

encountered in video compression. Typically, non-linear encoding is performed by a power function, say $\lambda < 1.0$ applied to the input values such that more bits are allocated to lower magnitudes where perceptual differences are more visible thus minimising the quantisation errors.

To encode chroma information existing algorithms such as *hdrv* and *fraunhofer* encode the chroma channels using the procedure similar to LogLuv [Lar98]. Although, the proposed algorithm uses the IPT color space which introduces a degree of non-linearity during conversion from the LMS cone excitation space to IPT, direct scaling and discretisation of the chroma channels to 10-bit integer representation leads to rounding error based visible contouring artefacts. Therefore, a further non-linear encoding step to the P and T channels is introduced by deriving the most appropriate power value(s) which when applied to the chroma information minimises the quantisation errors during discretisation.

The EMF is an optimisation function which minimises the difference between discretised floating point values such that $\mathrm{P},\mathrm{T} \in [0,1023]$ and their nearest integer calculated via a floor operation. The power value λ is derived as follows:

Let λ be the power value to be used for non-linear encoding, n be the targeted bit depth (10 in this case), $P_{inp} \in (0,1]$ be the input channel and $P_{out} \in [0,1023]$ be the output discretised channel. The application of the power function can be formulated as in equation 7.9.

$$P_{out} = \left\lfloor (P_{inp})^{\lambda} \cdot (2^n - 1) \right\rfloor \tag{7.9}$$

where the power function λ is derived by a brute-force technique which replicates the quantisation and de-quantisation steps, evaluates different values of $\lambda \in (0,1]$ such that the difference between

10-bit scaled floating point values and its nearest integer representation is minimal as shown in equation 7.10.

$$\text{argmin}\left(\frac{1}{MN}\sum_{j=1}^{N}\sum_{i=1}^{M}\left|\left(\frac{\lfloor (P_{inp})^{\lambda}\cdot(2^n-1)\rfloor}{(2^n-1)}\right)^{\frac{1}{\lambda}}-P_{inp}\right|\right) \tag{7.10}$$

where M and N represent the horizontal and vertical resolution, respectively. Upon application of the power values to the chroma channels, the λ values applied to each chroma channel is then stored as metadata and used later during reconstruction.

7.2.5 Metadata information

As a result of frame normalisation, intensity scaling, chroma scaling and non-linear encoding of chroma channels, the proposed algorithm produces a metadata information containing the scaling information of the intensity channel, the minimum and maximum values of the chroma channels prior to scaling and finally the power values applied to each chroma channel. This data is then stored in the form of a look-up table (LUT) for each frame and the final LUT is stored as a secondary metadata stream. The LUT structure is given in Table 7.2.

FrameNo	ν	I_{scale}	P_{min}	P_{max}	T_{min}	T_{max}	λ_P	λ_T
00000	4658	4000	-0.567	0.892	-0.124	0.589	0.899	0.967
00001								
.....								
00149								

Table 7.2: Example metadata information look-up table.

Although the metadata is vital for accurate reconstruction of HDR frames, the reliability on auxiliary information is not always desired for compression and transmission purposes as corruption of the metadata information would lead to faulty HDR reconstruction. Therefore, an alternate solution which eliminates the requirement for auxiliary metadata albeit at the cost of reconstruction quality is also proposed herein. To that end, a few constants need to assumed which are as follows:

- The intensity channel values are to be scaled such that $\mathrm{I}' \in [10^{-5}, 10^{4}]$ irrespective of the PTF applied to map the values to a 10-bit JND scale.
- The accurate scaling of chroma channel pixel values where $P, T \in [-1, 1]$ are to be replaced by a straightforward *addition and multiplication* routine in order to map them to a $[0, 1]$ range.
- The chroma error minimisation function is replaced with a fixed non-linearity based gamma encoding such that $\gamma = \frac{1}{2.2}$.

It should be noted that upon objective evaluation of the proposed algorithm with the assumed constants, the image reconstruction quality was slightly lower compared to the metadata solution as presented in the main manuscript.

7.3 Evaluation of compression algorithms

The compression performance of the proposed algorithm was evaluated against four state-of-the-art compression algorithms i.e. *pq*, *bbc-hlg*, *fraunhofer* and *hdrv* (see Section 3.3 for details), using 39 HDR video sequences across a range of energy-difference, structural and perceptual QA metrics. This section briefly discusses the evaluation methodology and the materials required to conduct the objective evaluation.

7.3.1 Materials

The materials used for this evaluation were the five compression algorithms including the proposed, the 39 HDR video sequences which represent a wide variety of scenes and overall dynamic range, seven QA metrics including the perceptual QA metrics and the x265 [Orgb] video codec.

7.3.2 Evaluation methodology

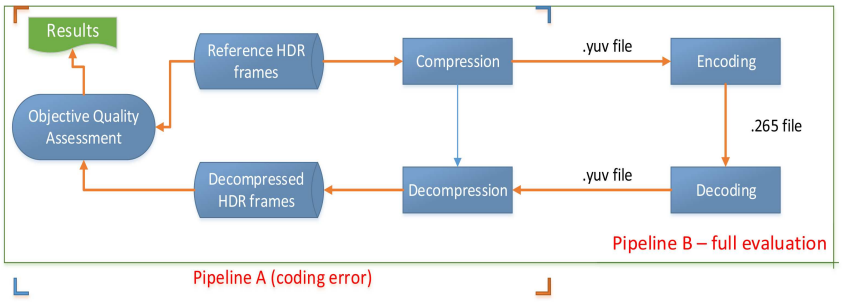


Figure 7.5: Schematic diagram of the evaluation methodology

The methodology can be classified into two parts. In *Pipeline A* (see Figure 7.5), the reference HDR frames from each of the 39 sequences were compressed using the five algorithms creating intermediate codec suitable files (labeled as HDRVs). Subsequently, the HDRVs are decompressed using the decompression part of the algorithms to reconstruct the HDR frames. The reference and reconstructed HDR frames are then evaluated using the objective QA metrics. In video compression, the results obtained by such an exercise computes the coding errors produced by each algorithm which determines compression quality without the external influence of the codec.

Pipeline B, extends *Pipeline A* and introduces the x265 codec. The HDRVs are passed to the codec which encodes the frames into a raw video stream which is subsequently

decoded and decompressed to reconstruct the HDR frames.

For a comprehensive evaluation of the algorithms at different quality levels, 150 frames from each of the 39 sequences were compressed using the five algorithms producing HDRVs which were subsequently 4:2:0 sub-sampled and then encoded at 11 different quality levels by controlling the quantisation parameter (QP) of the codec. The QP values were set such that $QP \in [0,5,10,...,50]$ where $QP = 0$ represents lossless encoding and $QP = 50$ represents a highly lossy compression. The group of pictures (GOP) sequence was **I-B-B-B-P** with an intra-frame period of 30.

The reference and reconstructed HDR frames were evaluated against a set of QA metrics and results obtained are first averaged over the number of frames (per sequence) followed by a cumulative average over 39 sequences. The averaged results are then used to plot the mean Rate Distortion (RD) graphs which exhibit the overall performance of the algorithms. However, the mean RD graphs do not provide the in-depth understanding since there is a significant amount of variation in both image reconstruction quality and bitrate required to encode the sequences depending upon the scene content. Therefore, in addition to the mean RD graphs, the evaluation data was used to plot interpolated RD graphs which exhibits the following:

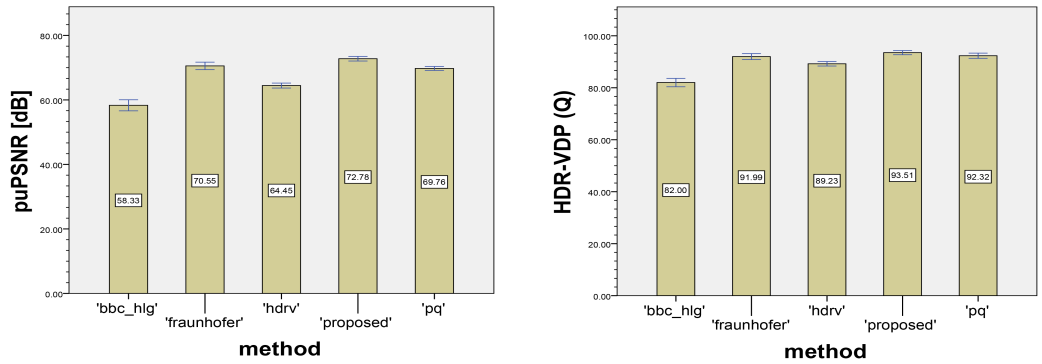
1. Variation in image reconstruction quality at fixed output bitrates $\in [0.2,2]$ bpp such that the range of bitrates reflect typical transmission bandwidth available in low to high end networks.

2. Variation in encoding at fixed image quality levels where reconstruction quality is of utmost importance.

A combination of the three sets of results provides a comprehensive understanding of the RD characteristics of each algorithm thereby allowing a fairer judgement of the proposed algorithm's performance against existing state-of-the-art.

7.4 Results

This section presents multiple sets of results obtained from the objective evaluation using a set of seven QA metrics described earlier in Chapter 4. The same set of error metrics were used for the objective evaluation described earlier in Chapter 6. Thus, following the evaluation methodology described in Section 7.3.2, the first set of results as shown in Figure 7.6 exhibit the coding error of each algorithm averaged over 39 sequences. It is to be noted that only the coding error results obtained from perceptual QA metrics such as puPSNR and HDR-VDP are presented here since the prediction by these error metrics have the highest correlation with subjective evaluation and are most likely to be noticed by the HVS [MDBR*16a].



(a) Coding error pu-PSNR (higher is better)

(b) Coding error HDR-VDP(Q) (higher is better)

Figure 7.6: Coding error of five algorithms - averaged over 39 sequences along with 95% confidence interval bars.

Next, as mentioned in Section 7.2 and described visually in Figure 7.3, the proposed algorithm is also a generic framework where several PTFs (see Section 7.1.2) can be used to perceptually encode the *intensity* channel. Since Figure 7.6 demonstrates the coding error of the proposed algorithm using the novel hybrid PTF, the performance of the proposed hybrid PTF should be compared against existing PTFs to determine its suitability for perceptual encoding of intensity values. Such a comparison however can only be conducted upon the introduction of the codec. Therefore, following *Pipeline B*), the five PTFs (including the proposed hybrid PTF) were used in conjunction with the algorithm to compress the 39 HDR video sequences and evaluated against the set of QA metrics. Figure 7.7 presents the mean as well interpolated RD characteristics of the proposed algorithm when used in conjuction with each of the five PTFs as measured by both perceptual and structural QA metrics.

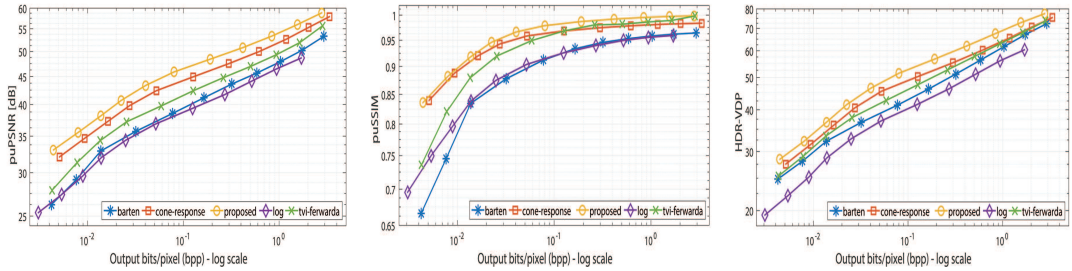
With the performance of the five PTFs when used in conjuction with the proposed algorithm established, Figures 7.8, 7.9 and 7.10 present a comprehensive set of RD characteristics where the proposed algorithm (using the hybrid PTF) has been evaluated against the four existing state-of-art solutions using seven QA metrics.

7.5 Discussion

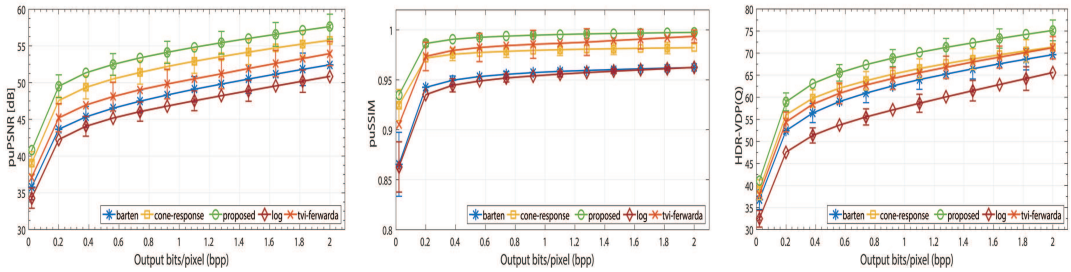
This section combines all the results shown in Section 7.4 and provides an in-depth performance analysis of the proposed algorithm. The analysis in this section can be subdivided into three different discussions as given below:

7.5.1 Coding errors

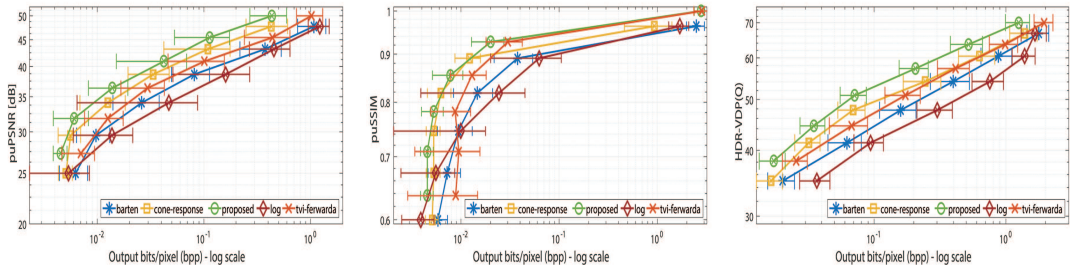
Figure 7.6 in Section 7.4 demonstrates the mean coding error of each algorithm averaged over 39 HDR video sequences. Based on the image reconstruction quality as measured by puPSNR and HDR-VDP, the mean and variation of coding errors demonstrate that the



(a) Perceptual RD characteristics of five PTFs - puPSNR (b) Structural similarity index - puSSIM (c) Perceptual RD characteristics five PTFs - HDR-VDP(Q)



(d) Perceptual image quality variation - puPSNR (e) Structual similarity variation puSSIM (f) Perceptual image quality variation - HDR-VDP(Q)



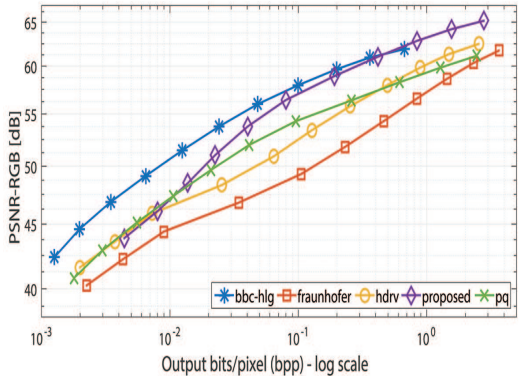
(g) Encoding bitrate variation - puPSNR (h) Bitrate variation - puSSIM (i) Bitrate variation - HDR-VDP(Q)

Figure 7.7: Mean and interpolated RD characteristics of the proposed algorithm with five different PTFs - averaged over 39 sequences (interpolated data exhibits variation with 95% confidence interval).

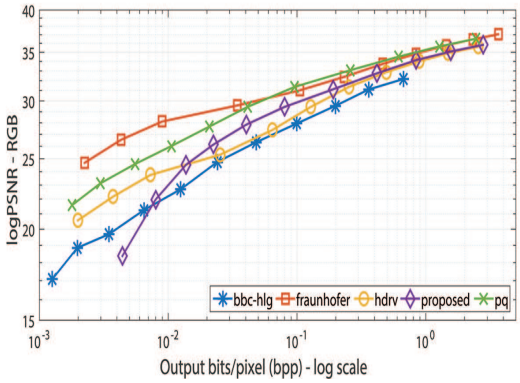
proposed algorithm exhibits less coding error than existing solutions. Amongst the existing solutions, the best performance is exhibited by *pq* and *fraunhofer* while *bbc-hlg* exhibits the least desired performance. The variation in coding errors is relatively low for all algorithms.

7.5.2 RD characteristics of the five PTFs

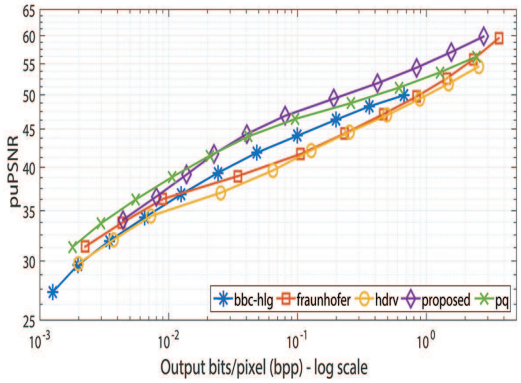
Next, the mean and interpolated results in Figure 7.7 suggest that the proposed hybrid PTF outperforms existing PTFs both in terms of image reconstruction quality and encoding bitrate. The mean RD characteristics shown in Figures 7.7a, 7.7b and 7.7c demonstrate that amongst the established PTFs, GCRM produces the best encoding results while the least desired performance is exhibited by the adaptive logarithmic TF. This can be attributed to the fact that GCRM performs a conservative quantisation of a large portion of the scaled inten-



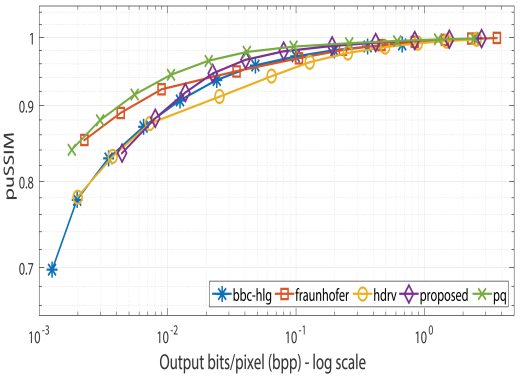
(a) RD characteristics PSNR-RGB (higher is better)



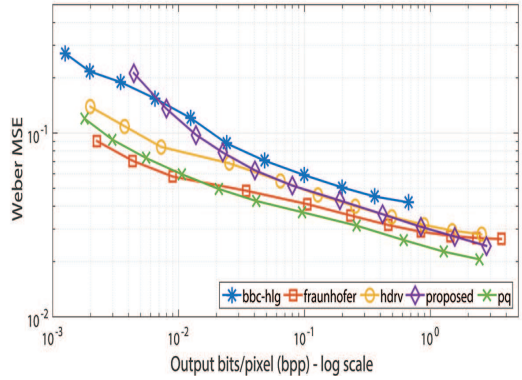
(b) logPSNR-RGB (higher is better)



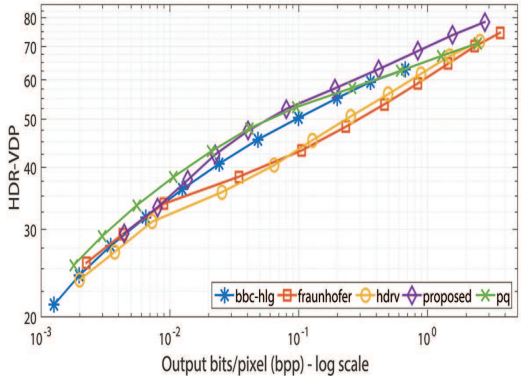
(c) puPSNR (higher is better)



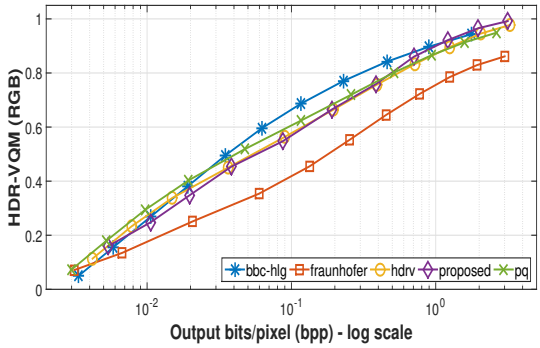
(d) puSSIM (higher is better)



(e) Weber

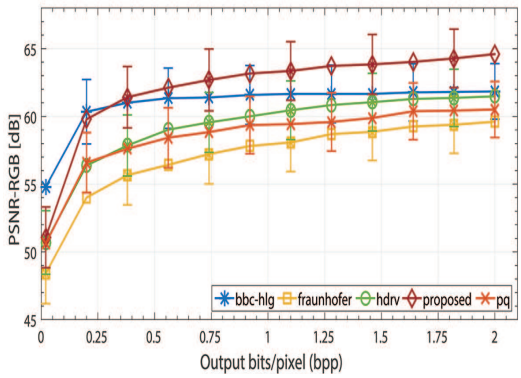


her is better)

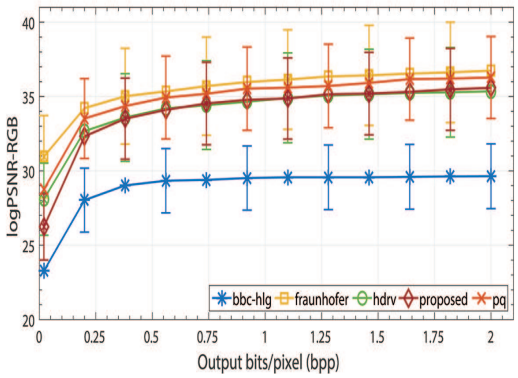


(g) HDR-VQM (higher is better)

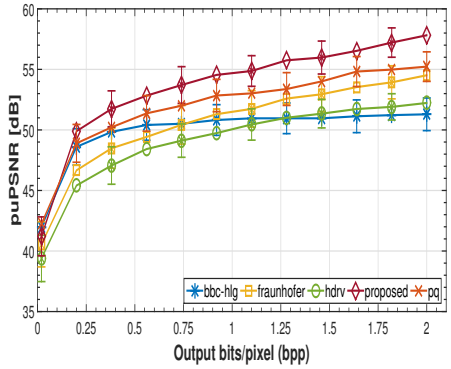
Figure 7.8: Mean RD characteristics of the five algorithms - averaged over 39 sequences across seven QA metrics.



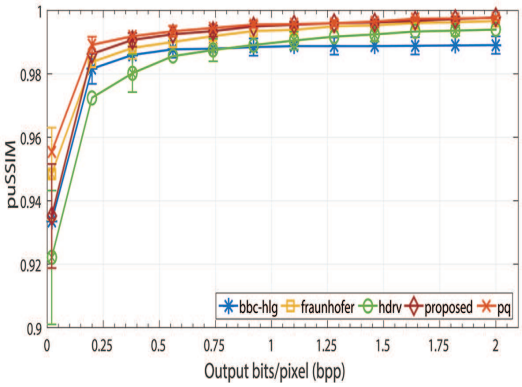
(a) RD characteristics PSNR-RGB (higher is better)



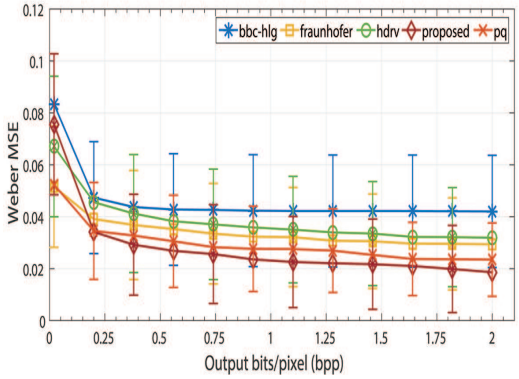
(b) logPSNR-RGB (higher is better)



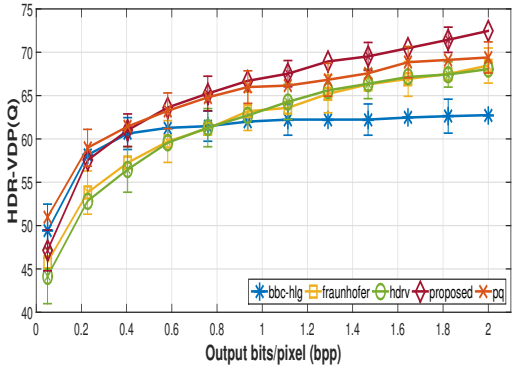
(c) puPSNR (higher is better)



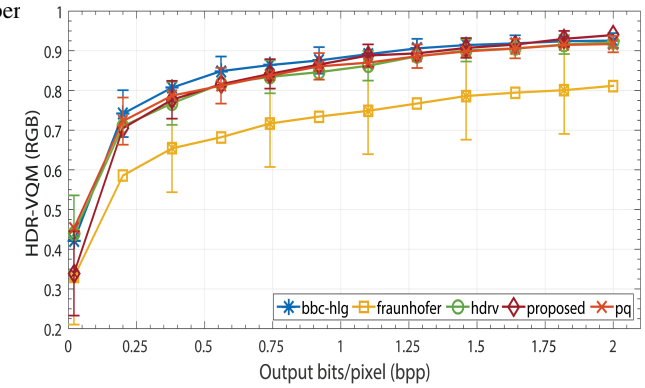
(d) puSSIM (higher is better)



(e) Weber

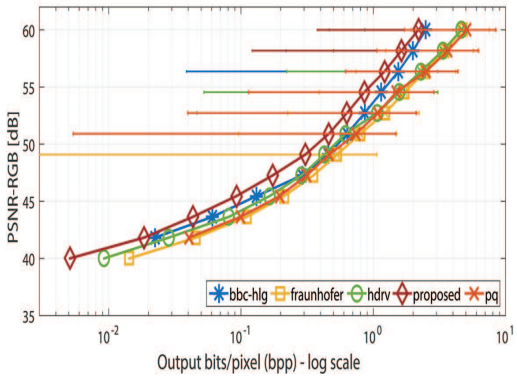


igher is better)

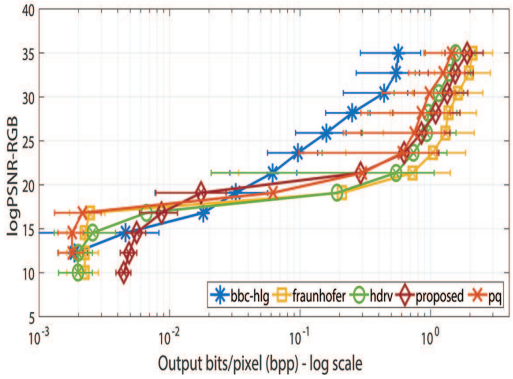


(g) HDR-VQM (RGB) (higher is better)

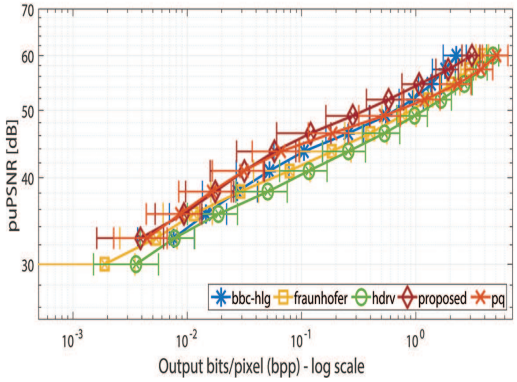
Figure 7.9: Interpolated RD characteristics of the five algorithms at fixed bitrates (exhibiting variation in image quality) - averaged over 39 sequences.



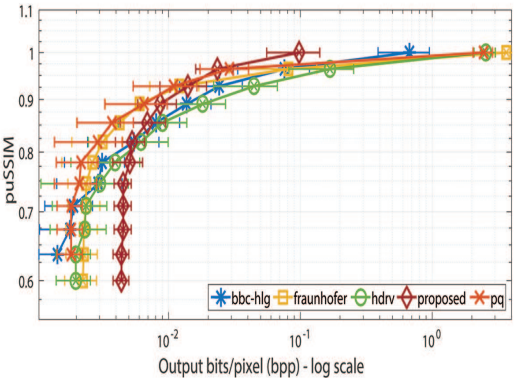
(a) RD characteristics PSNR-RGB (higher is better)



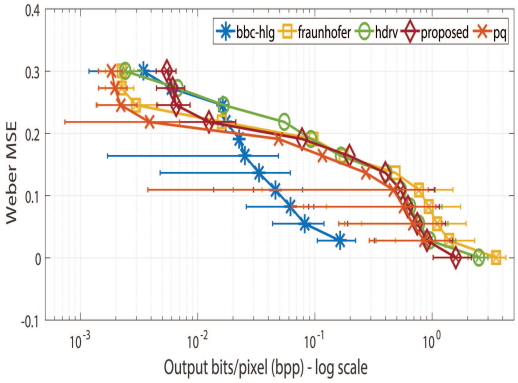
(b) logPSNR-RGB (higher is better)



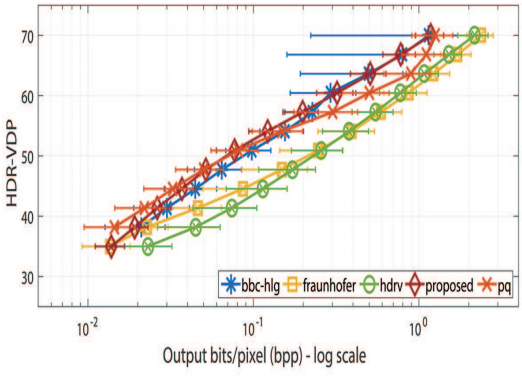
(c) puPSNR (higher is better)



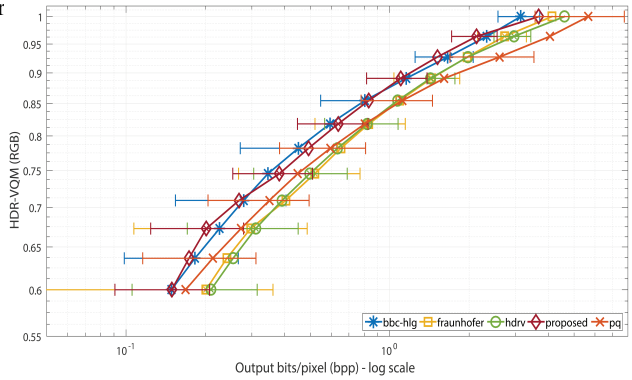
(d) puSSIM (higher is better)



(e) Weber



her is better)



(g) HDR-VQM (higher is better)

Figure 7.10: Interpolated RD characteristics of the five algorithms at fixed quality levels (exhibiting variation in bitrate) - averaged over 39 sequences.

sity values while allowing a coarser quantisation of the darker regions [SYD87, MMS06]. In terms of structural similarity during reconstruction, Figure 7.7b suggests that the proposed hybrid PTF outperforms both GCRM and Ferwarda's *t.v.i* by being able to facilitate the best structural reconstruction of HDR frames. This can be attributed to the fact that the hybrid PTF mitigates the shortcomings of GCRM and facilitates a finer quantisation in the darker regions as well as maintaining the quantisation exhibited by Ferwarda's *t.v.i* and/or Barten's PTF for high intensity regions.

The mean results are consistent with the interpolated results given in Figures 7.7d, 7.7e and 7.7f which presents a clearer reflection of the mean and variation in image reconstruction quality as exhibited by each PTF. The results suggest a straightforward ordinal ranking where the hybrid PTF followed by GCRM and Ferwarda's *t.v.i* are the three top performing PTFs.

In addition to the image reconstruction quality, the results demonstrated in Figures 7.7g and 7.7i also suggest that the hybrid PTF not only facilitates better image reconstruction but also requires less bandwidth/transmission cost to delivery high-fidelity HDR video reconstruction. Furthermore, the proposed PTF provides an easy to implement analytical solution to map scaled *intensity* channel values to 10-bit JND scaled luma space. Finally, although the proposed PTF exhibits superior performance, the perceptual difference with established PTFs such as GCRM and Ferwarda's *t.v.i* is minor. Therefore, both GCRM and the modified Ferwarda's *t.v.i* can be used in conjunction with the proposed algorithm albeit with minor quality degradation.

7.5.3 Evaluation results

This section analyses the RD characteristics obtained by evaluating the proposed algorithm (in conjunction with the proposed PTF) against the existing solutions using seven QA metrics. The mean RD characteristics are given in Figure 7.8 while the interpolated RD characteristics are given in Figures 7.9 and 7.10.

The first set of results demonstrate the overall RD characteristics of the five compression algorithms at 11 different quality levels averaged over 39 HDR video sequences as shown in Figure 7.8. The results obtained from the perceptual and structural metrics such as puPSNR, puSSIM and HDR-VDP exhibit that the proposed algorithm outperforms the four existing algorithms. Also, there is a high correlation between Figures 7.8c and 7.8f and it has been seen that the results obtained puPSNR and HDR-VDP tend to have high-very correlation with subjective evaluation [MDBR*16a]. On the other hand, the results obtained from other energy difference metrics such as PSNR, logPSNR and Weber MSE exhibit that some of the existing solutions perform at par or better than the proposed algorithm. However, previous research [MDBR*16a, ABDD*14, HRE15, HBP*15] have concluded that the correlation between these QA metrics and subjective evaluation is significantly lower compared to perceptual QA metrics . An exception to this are the RD characteristics

demonstrated in the HDR-VQM results (Figure 7.8g) where *bbc-hlg* outperforms the rest of the four algorithms except at very high bitrates. Also, the image reconstruction quality of *fraunhofer* is significantly lower in HDR-VQM, a phenomenon not reflected by the other QA metrics.

The second set of RD graphs as shown in Figure 7.9 provide a clearer reflection of the compression performance. Here, the PSNR and Weber MSE results in Figures 7.9a and 7.9e has a positive correlation with the puPSNR and HDR-VDP results shown in the main manuscript. The only exception to these results are that obtained from the logPSNR metric. However, as discussed in [MDBR*16a], logPSNR does not exhibit high correlation with subjective evaluation.

The third set of results presented in Figure 7.10 exhibit the variation in encoding bitrates at fixed quality levels. Such a comparison is a fairer judgment of the algorithms' compression performance. Figures 7.10c, 7.10f and 7.10g reveal that the proposed algorithm is capable of reconstructing high fidelity HDR video at lower transmission cost than existing algorithms. However, the Figure 7.10g also reflect that the *bbc-hlg* algorithm performs at par or marginally better than the proposed algorithm in terms of transmission cost. Interestingly, the puSSIM results are not in agreement with the other perceptual QA metrics. However, the PSNR results (see Figure 7.10a) show a positive correlation with the results reflected by some of the perceptual QA metrics.

Finally, as a general observation it can be seen that the RD graphs (especially the perceptual QA metrics) in Figures 7.8, 7.10 and 7.9 clearly demonstrate that the performance of *hdrv* and *fraunhofer* is comparable analogous to the results shown in [MDBR*16a]. However, the performance of *pq* and *bbc-hlg* are not comparable with *pq* demonstrating the best performance amongst the existing solutions.

7.6 Conclusion and Future Work

To summarise and conclude this work proposes a novel *non-backward* compatible HDR video compression algorithm along with a novel hybrid PTF and a non-linear chroma error minimisation function. Comparative results suggest that the proposed PTF facilitates better preservation of scaled intensity values than established PTFs. Furthermore, the combination of IPT colour space, proposed PTF and the error minimisation function is not only able to deliver better HDR image reconstruction but requires lower bitrates than existing solutions.

To summsarise, the adoption possibility of the proposed algorithm can be considered. Any new HDR video compression algorithm has to satisfy the following properties a) the algorithm should be able to perform better than existing solutions on uncompressed data, b) it should be able to provide better image reconstruction at lower transmission cost and c) add new functionality hitherto unavailable. The proposed algorithm (in the form of the framework) satisfies all three properties. It exhibits lower coding errors than existing

solutions, it facilitates better image reconstruction at lower transmission costs compared to existing solutions, the design of algorithm in the form of a framework allows the plug-and-play capability to map scaled *intensity* values to 10-bit JND space using established PTFs in addition to the proposed PTF and finally the algorithm implements a novel chroma encoding technique which has largely been ignored to date.

Future work in this area could include the possibility of designing a PTF which further reduces the quantisation errors and provides better reconstruction capability of compressed HDR video. Furthermore, the proposed algorithm can also be further evaluated across other HDR video sequences especially those processed with the BT.2020 primaries [Rec12] and higher bit-depths such as 12- and 14-bits/pixel/channel as and when optimised codec support becomes available.

7.7 Summary of the design decisions

This section provides the summary of the design decisions and parameters taken into account in order to design the proposed algorithm.

- The usage of colour opponent space IPT (over established CIELAB/CIELUV) provides a better decorrelation of the achromatic and chromatic information. It also eliminates the hue compression issues of CIELAB and CIELUV.

- A perception based analytical transfer function provides straightforward encoding of scaled *intensity* channel to JND spaced luma code values. Evaluation of PTFs using contrast vs. intensity curve provides further refinement of the proposed PTF not only to eliminate contrast jumps but also for optimal allocation of bit codes values. This also provides a detailed understanding of the effects and encoding efficiency of different PTFs such as GCRM, Ferwarda's *t.v.i* and GDF.

- The effect of non-linear error minimisation function for accurate retention of chroma information compared to existing state-of-the-art.

- The effect of metadata information in accurate reconstructing HDR video frames. Section 7.2.5 also describes a number of techniques to eliminate the use of metadata information albeit at the loss of reconstruction quality.

Chapter 8

Conclusion

THIS thesis has introduced a novel HDR video compression algorithm which attempts to deliver better HDR video reconstruction at lower storage/transmission costs compared to existing state-of-the-art. In the development of this novel HDR video compression algorithm, this thesis studies the various design decisions and parameters required to create efficient HDR video compression algorithms. As a preliminary step to answer the research question, Chapter 5 validated the user preference of HDR video content over LDR video content. Chapter 6, conducted an objective and subjective evaluation of existing HDR video compression algorithms and also introduced a robust evaluation methodology which can be be used for such future evaluations. Finally, Chapter 7 introduced the novel HDR video compression algorithm and evaluated this against existing solutions. This chapter summarises the contributions, draws conclusions and provides an outlook on future work.

8.1 Preliminary verification

A number of challenges in every aspect of the HDR pipeline (capture, storage, processing and display) still remain and need to be solved before HDR can be fully adopted in mainstream media. A fundamental question emerges as to whether that investment is worth the effort if the inherent advantages of HDR over LDR are imperceptible by naïve users. To date, very little research has been conducted to explore the practical feasibility and acceptability of HDR over existing LDR and this was primarily targeted for static images. No such work existed for HDR videos.

Chapter 5, explored the acceptability of HDR videos over existing LDR videos purely from a viewers' perspective. Six HDR video sequences based on their overall dynamic range, out of a repository of 39 sequences. Three state-of-the-art HDR to LDR mapping techniques to were used to generate their LDR counterparts. A rating- and a ranking-based subjective experiment was conducted with 28 and 27 users (after outlier removal), respectively. The users were tasked to rate and rank the candidate sequences based on their preference.

The overall result from the rating experiment demonstrated that on a rating scale of $R_{preference} \in [0, 10]$, where higher is better, HDR video representations R_{HDR} were rated at ≈ 7.10 with 95% confidence interval bounds. The LDR counterparts R_{LDR} were rated at ≈ 6 which is $\approx 9\%$ lower. Although, from the averaged rating scores, this might not seem to be a significant difference, further analysis indicated that amongst the four representations of an HDR sequence (one HDR and three LDR), a statistically significant difference exists between the HDR representation and its LDR counterparts. Amongst, the three LDR representations, the image appearance model *icam* (rated at ≈ 6.52) is preferred over the TMO *mantiuk* representation and exposure extraction technique *optimal*.

The ranking results demonstrated that on a ranking scale where $R_{preference} \in [1, 4]$ (lower is better), the average ranking of HDR video sequences R_{HDR} were ≈ 1.54 with 95% confidence interval bounds. The average ranking of the LDR counterparts were ≈ 3 which is $\approx 36\%$ lower than the HDR representations. Further analysis of the ranking results revealed that there exists a statistically significant difference between the HDR and LDR representations although no such difference were found in-between the LDR representations.

A detailed methodology was provided for conducting subjective experiments required for such an evaluation including a methodology for analysis of the subjective results. The key conclusions that can be drawn are:

1. Given the right viewing conditions, viewers prefer the HDR video representation of a scene over its LDR counterparts.

2. HDR to LDR mapping techniques including state-of-the-art TMOs, image appearance models and exposure extraction techniques are unable to deliver the details, scene reproduction capability and the immersive experience provided by HDR.

There are, however some limitations to this study. For instance, only six HDR sequences were used in this evaluation out of which only five were used for the ranking-based experiment. The results presented in Chapter 5 might vary if the number of sequences and participants are increased. Furthermore, the viewers were presented with independent visual stimuli which were not a part of any contextual narrative (such as a short-film) upon which the results might also vary.

8.2 Evaluation of existing HDR video compression algorithms

Following the key conclusions of Chapter 5, Chapter 6 attempts to gain an in-depth knowledge of existing HDR video compression algorithms and more specifically their design aspects which include a thorough understanding of the following:

1. The schematic advantages and disadvantages of the *non-backward* compatible and *backward* compatible algorithms.

2. Knowledge of several colour space transformations and perceptual transfer functions required to effectively manipulate HDR pixel values for HDR video compression purposes.

3. The chroma preservation techniques used in HDR video compression.

4. The effect of noise reduction and other data manipulation steps to reduce output file size.

5. The basic working principles of video codecs including the effect of higher bit-depth encoding.

In addition to the mentioned design aspects, Chapter 6 also provided a thorough understanding of the objective and subjective quality assessment techniques required to evaluate existing and future HDR image/video compression algorithms.

Six state-of-the-art HDR video compression algorithms were implemented and an objective evaluation was conducted using a set of 39 HDR video sequences and seven dedicated/modified HDR QA metrics. Results obtained from the objective evaluation as plotted by the mean and interpolated rate-distortion graphs demonstrate that while *non-backward* compatible algorithms such as *hdrv* and *fraunhofer* are able to deliver high-fidelity HDR video (HDR-VDP(Q) ≈ 70, HDR-VQM ≈ 0.98) at output bitrates of $\approx 1-3$ bits/pixel (bpp), *backward* compatible algorithms such *hdrjpeg*, *hdrmpeg* and *gohdr* are only able to deliver similar reconstruction quality at bitrates $\geq 7-8$ bpp thus indicating significantly higher storage/transmission costs. A similar objective evaluation with six short-listed HDR video sequences also demonstrate that *hdrv* and *fraunhofer* are able to deliver similar reconstruction quality at output bitrates of ≈ 1.0 bpp while *backward* compatible algorithms deliver the same quality at bitrates ≥ 4.0 bpp.

In addition to the objective evaluation, two ranking-based subjective experiments were conducted with six short-listed sequences by 30 mutually exclusive group of participants at two fixed quality levels. The combined subjective results (from both experiments) suggested that on a ranking scale $R \in [1,7]$ (lower is better), *hdrv* and *fraunhofer* received an overall ranking of 2.62 and 3.02, respectively. *Backward* compatible algorithms such as *hdrjpeg* and *hdrmpeg* received an overall ranking of 4.90 and 5.22, respectively. Also, the combined Kendall's coefficient of concordance amongst the participants was $W = 0.597$ which indicates a high degree of agreement amongst the participants.

The combined objective and subjective results demonstrate that while the *non-backward* compatible algorithms *hdrv* outperformed all existing solutions, the *backward* compatible algorithm *rate* exhibited the least desired performance amongst the six chosen algorithms. In addition to the main results, a correlation between the objective and subjective results revealed that modified/dedicated perceptual QA metrics such as puPSNR, puSSIM, HDR-VDP and HDR-VQM had a high-very high correlation $(0.8-1.0)$ with subjective results and could thus accurately indicate/predict the performance of the algorithms

in question. Also, there exists a high correlation in-between the perceptual QA metrics. However, the correlation reveals that traditional energy-difference QA metrics (even though modified for HDR) are less appropriate indicators of HDR video reconstruction quality (correlation ≈ 0.37). An indirect inference which can be drawn from the evaluation is that any new algorithm which performs well against perceptual QA metrics is also likely to perform well in subjective evaluation thereby ensuring the acceptability of the new algorithm.

The detailed evaluation and analysis provided a thorough understanding of the several design aspects and HDR data manipulation steps required for high-fidelity HDR video reconstruction. Chapter 6 also proposed a detailed methodology based on which other HDR video compression evaluations can be conducted in the future.

Limitations of this study are:

1. All compression algorithms selected for this work were proposed before the MPEG CfE [LFH15]. Therefore, this study does not include the recent proposals made to MPEG.

2. The selected compression algorithms were implemented from the original papers and although all efforts were made to verify most of the implementations (with the original authors), it is not possible to guarantee complete accuracy of the re-implementations and

3. Finally, due to time constraints, only two subjective experiments at two different quality levels were conducted.

8.3 Uniform colour space based novel HDR video compression algorithm

The knowledge gained from the evaluation in Chapter 6 revealed the advantages and disadvantages of state-of-the-art HDR video compression algorithms. It also provided an understanding of several design aspects of these algorithms such as colour space transformations, use of perceptual transfer functions, use of residual luminance information, auxiliary metadata information and the basic working principles of the H.264/AVC video codec along with its limitations. Furthermore, the literature on recent HDR video compression proposals to MPEG CfE provided meaningful insights about the efficient perceptual and mathematical optimisation techniques required to preserve luminance and chroma details for compression purposes.

The combined knowledge from the evaluation and existing literature was instrumental in the design of a novel HDR video compression algorithm. This is introduced in Chapter 7. The salient features of the proposed algorithm are that is it exploits:

1. A state-of-the-art perceptually uniform colour opponent spaces such as IPT for optimal decorrelation and effective manipulation of the achromatic and chromatic information.

2. An optimised analytical perceptual transfer function for effective JND encoding of the *intensity* channel. The transfer function was optimised using a *c.v.i* curve for more efficient allocation of targeted bit-depth while enforcing C^1 continuity.

3. An optimisation technique (error minimisation function - EMF) to accurately preserve chroma information.

Also, in addition to the novel PTF, the proposed algorithm also acts as a framework which enables the use of other existing PTFs to encode the *intensity* channel.

Chapter 7 justifies the usage of the IPT colour opponent space as opposed to CIELAB/CIELUV. It provides an overview of the advantages and disadvantages of four widely used PTFs which can be used to map the scaled *intensity* channel to JND spaced luma code values. Based on the advantages and disadvantages of the existing PTFs, Chapter 7 also introduced a novel PTF with an analytical solution which was further optimised using a *c.v.i* curve. An EMF to accurately preserve chroma information was also proposed as a part of this algorithm. The proposed EMF performs a non-linear encoding of the chroma information, similar to gamma encoding albeit with higher precision to minimise the difference between floating point and discretised integer representation of the pixel values.

The proposed algorithm was evaluated in two stages using the same objective evaluation methodology described in Chapter 6. They are:

1. First, the *intensity* channel encoding efficiency of the five PTFs (including the proposed PTF) when used in conjunction with the rest of the algorithm were compared using rate-distortion plots. Perceptual QA metric results suggested that the image reconstruction quality of the proposed algorithm improved by $\approx 8-10\%$ while using the proposed PTF as opposed to existing PTFs. Also, the interpolated rate-distortion plots at fixed quality levels show that the algorithm was able to achieve $\approx 15-20\%$ bitrate savings when using the proposed PTF. GCRM and Ferwarda's *t.v.i* exhibited the best performance amongst the existing PTFs.

2. Second, the proposed algorithm along with the proposed PTF and EMF was evaluated against four state-of-the-art *non-backward* compatible algorithms to determine the following:

    - The coding error of the proposed algorithm compared to existing solutions without the influence of the video codec. Although all five algorithms exhibited high performance figures, puPSNR and HDR-VDP results suggested that the

proposed algorithm outperformed *pq* (the best performing existing solution) by $\approx 4.32\%$ and $\approx 1.38\%$, respectively.

- The overall performance (output bitrate versus reconstruction quality) of the algorithm at 11 different quality settings (video codec) plotted with mean and interpolated rate-distortion graphs. Results obtained from the mean and interpolated rate-distortion plots demonstrated that the proposed algorithm outperformed the existing solutions at output bitrates ≥ 0.4 bpp. Also, overall results obtained from the puPSNR, HDR-VDP and HDR-VQM plots suggested that the proposed algorithm outperformed *pq* by approximately $6-8\%$ at comparable output bitrates.

Although, the proposed algorithm outperforms state-of-the-art existing HDR video compression algorithms, there are limitations to this work. They are as follows:

1. The algorithm's computation time is higher compared to the four existing solutions. This can be attributed to the fact that the proposed algorithm has a more complicated colour space transformation procedure than existing solutions. Also, the proposed EMF to optimise chroma preservation is a brute-force optimisation technique which searches for the most appropriate power value $\lambda \in (0,1)$.

2. The proposed PTF has only been evaluated against four existing PTFs. Results might vary with the introduction of other existing PTFs.

3. The proposed algorithm has been evaluated against four existing algorithms. Results might vary with the introduction of other *non-backward* compatible algorithms. Also, the proposed algorithm requires further objective and subjective evaluations with more HDR video sequences.

8.4 Summary of the design decisions

The primary motivation of this thesis was to explore the design decisions required to deliver high-fidelity HDR video at minimal storage and/or transmission costs. Summarising the knowledge gained from literature as well as an in-depth understanding of the crucial aspects HDR video compression from Chapter 6 and Chapter 7 it can be inferred that there are multiple factors which need to be considered while designing an HDR video compression algorithm that delivers high-fidelity HDR video at minimal storage/transmission costs. These are as follows:

1. Choice of compression approach. Chapter 6 clearly shows that *non-backward* compatible algorithms deliver better quality at lower bitrates than *backward* compatible algorithms. However, *backward* compatible algorithms can be used by legacy video infrastructure.

2. Colour space transformations for effective de-correlation of achromatic and chromatic information. Perceptually uniform colour spaces that are considered are CIELAB,CIELUV, or IPT.

3. Preservation of achromatic information using the most appropriate perceptual/opto-electronic transfer functions. Verification of this is done using $c.v.i$ plots.

4. Linear/non-linear chromatic information preservation using optimisation functions/gamma encoding.

5. Use of metadata to effectively reconstruct HDR video frames.

Another indirect yet relevant factor is the choice of video codec since state-of-the-art codecs such as the HEVC (including the optimised implementation x265) are able to provide up to 40% bitrate savings compared to the previous generation of codecs such as H.264/AVC while achieving the same image reconstruction quality [BDAPN14].

The primary research question introduced in Chapter 1 has been answered by the algorithm proposed in Chapter 7. This takes the mentioned design factors into consideration and is able to deliver $\approx 6-8\%$ superior HDR video reconstruction quality at approximately 15% lower storage/transmission cost.

8.5 Future work

HDR video compression is a relatively new field of research and there remains a multitude of interesting research opportunities as well as many pending issues which need to be mitigated before HDR video can be commercially introduced in mainstream media.

A number of efficient HDR video compression algorithms have been proposed to date and some have even been standardised such as the SMPTE 2084 standard [PQ14] and HLG (ARIB-B67) [ari15]. However, no one standardised compression algorithm has been chosen to date. Therefore, comprehensive objective and subjective evaluations with other algorithms using more sequences are still required to select the best performing compression algorithm.

Another important aspect is the video codec. Although the HEVC codec along with the efforts of the MPEG committee has made remarkable progress in encoding HDR video content, it is still under active development and unable to provide a 4:4:4 sub-sampled encoding of 16 bit HDR video content. Therefore, further development and comprehensive testing is required for native support of HDR video content. Finally, displaying HDR content on native HDR displays is another challenge. Even though prototype display devices with a peak luminance of 4000 $\mathrm{cd/m^2}$ [SIMa] and ≥ 6000 $\mathrm{cd/m^2}$ [SIMb] are being built, these devices these devices require further refinement prior to commercial adoption. One of the major challenges is to reduce the amount of heat generated by these displays which

inhibits their longevity. Therefore, it is evident that even though HDR imagery has come a long way in the last decade, it is still not ready for commercial deployment unless the pending pipeline issues are mitigated.

8.5.1 Evaluation of *backward* compatible HDR video compression algorithms

Although the evidence presented in Chapters 6 and 7 clearly suggest that *non-backward* compatible algorithms provide enhanced image reconstruction quality at a much lower transmission cost compared to *backward compatible* algorithms, the codec suitability of such algorithms is an existing issue since higher bit-depth encoded files cannot be played back using legacy decoders and video players. Moreover, most hardware based encoder and decoder implementations are typically limited to 8-bits/pixel/channel. For initial adoption of HDR video,a comprehensive evaluation of *backward* compatible algorithms can be conducted (using a similar evaluation methodology as introduced in Chapter 6) to evaluate the best performing algorithm. Subsequently the chosen algorithm can be used as an intermediate step till dedicated higher bit-depth pipelines are ready for adoption.

8.5.2 HDR VQA metric

Another interesting area of research is the design, development and subsequent evaluation of a single dedicated HDR video QA metric which takes energy-difference, structural and perceptual errors into account. The evidence presented in Chapter 4, Chapter 6 and Chapter 7 suggest that modified/dedicated perception based HDR QA metric exhibit high to very high correlation with subjective results. However, perceptual accuracy is only a partial aspect of image quality evaluation and for a number of scientific applications, energy-difference and structural similarity are more important than perceptual accuracy. Also, most modified/dedicated perceptual QA metric such as HDR-VDP focus on the luminance similarity thereby ignoring colourimetric precision. Furthermore, only a few dedicated HDR video QA metric such as HDR-VQM and DRI-VQM exist to date. The absence of a standard video QA metric presents significant issues in HDR video quality evaluation, especially for compression related purposes. Design and development of a single QA metric which takes into account energy-difference, structural and perceptual aspects of image quality assessment in addition to spatio-temporal and colourimetric losses would be a valuable addition in this field. The structural similarity aspect can be taken into account using advanced signal processing techniques such as wavelet based decomposition and band-pass filtering whereas the perceptual and temporal aspects can be accounted for using perception based transfer functions and short/long term pooling of errors analogous to existing metrics such as HDR-VQM.

With the HDR pipeline issues successfully dealt with along with the development and standardisation of compression algorithms as well as quality assessment metrics, HDR

video techniques could finally be adopted in mainstream media to eventually phase out the existing LDR imagery.

8.6 Final remarks

HDR video techniques have taken concrete and progressive steps towards becoming a reality. It is now being actively used in many industries such as game design, media and entertainment, scientific and security. Although achieving an indistinguishable representation of the reality is still out of reach, the gradual improvement of capture, storage and display technologies increases the probability of replacing traditional imaging with HDR in the near future. The suggested design decisions as well as the novel algorithm proposed in this thesis is one step forward in solving the storage issue. Although *backward* compatible algorithms can still be used for early adoption of HDR, the widespread usage of HDR imagery is likely to only occur when some of the major issues in the HDR pipeline are solved and there is increased support of hardware- and/or software-based higher bit-depth video codecs. The contributions of this thesis are a step towards achieving accurate the light and colour reproduction of a scene.

Appendix A

A framework for HDR video evaluation

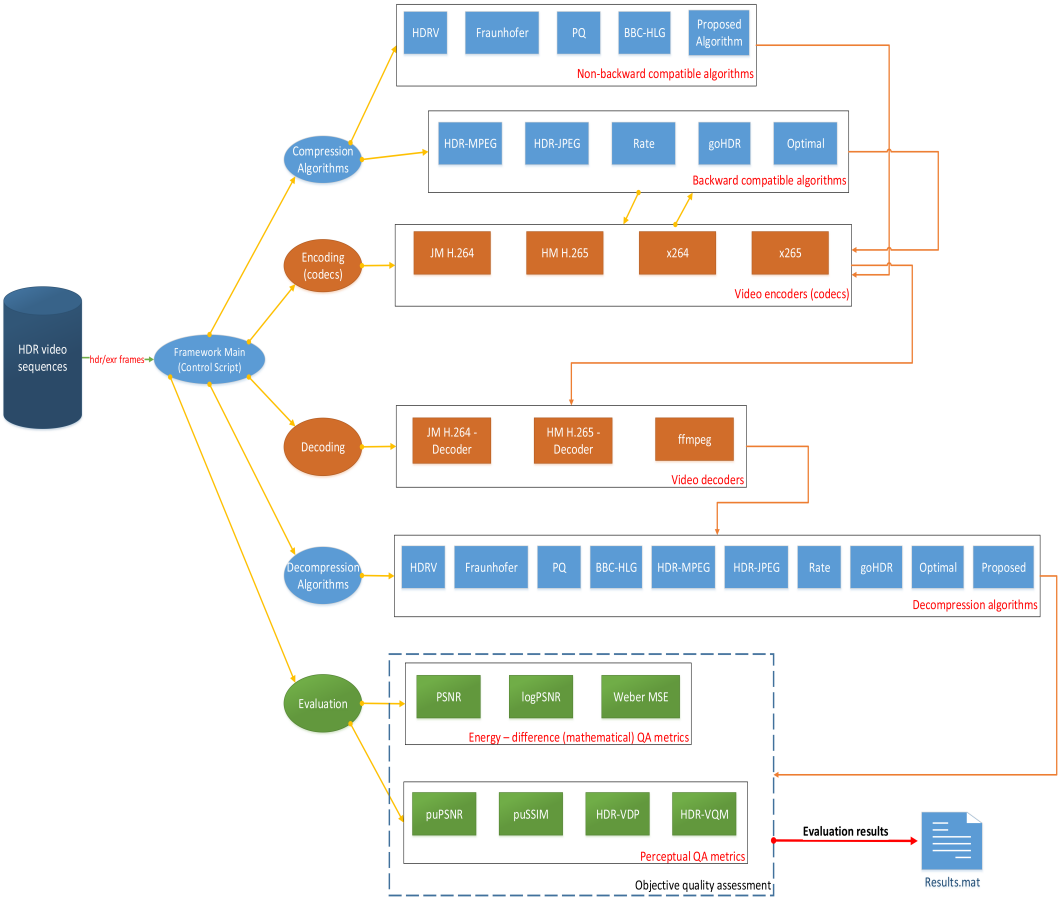


Figure A.1: Schematic diagram of the framework for HDR video quality evaluation

THIS chapter describes the framework that was created for objectively evaluating the HDR video compression algorithms as previously mentioned in Chapters 6 and 7.

A.1 Overview

As seen in Figure A.1, the framework has a modular structure and can broadly be divided into five modules as described below:

1. **Compression module:** Consists of the compression part of the HDR video algorithms which include non-backward and backward compatible algorithms. Further details are described in Section A.2.

2. **Video codec module:** This module consists of the several video codecs for 8-14 bits/pixel/channel video encoding. Further details are described in Section A.3.

3. **Video decoding module:** This module consists of several decoders to decode the encoded video streams. Further details are discussed in Section A.4.

4. **Decompression module:** This module consists of the decompression part of the HDR video compression algorithms which essentially reverses the compression process to reconstruct the HDR frames. Further details are described in Section A.5.

5. **Evaluation module:** This module consists of several objective QA metrics which include dynamic range dependent, dynamic range independent, structural and perceptual QA metrics for HDR image/video quality evaluation. Further details are described in Section A.6.

The most beneficial factor of the framework's modular structure is that compression algorithms, codecs and QA metrics can be added and removed with little or no modification to the framework structure. Following the addition, only the specific compression-decompression algorithm with specific codec can be used to evaluate the performance of the algorithm. Also, a combination of algorithms can be used with any combination of codecs and metrics for comprehensive evaluation. However, there are certain constraints which is discussed later in Section A.3.

A.2 Compression module

This module consists of the compression parts of the *non-backward* compatible algorithms such as *hdrv, fraunhofer, pq, bbc-hlg* (see Chapter 3 for details) and the proposed algorithm described in Chapter 7. The compression part of the HDR video compression algorithms convert input HDR frames to a codec suitable intermediate *'.yuv'* format which is then passed on to the codec module for encoding the *.yuv* files to encoded video stream. However, the bit-depth requirements of the algorithm are also taken into account to choose the type of the codec used for the encoding the *.yuv* files. Section A.3 describes this issue in more detail.

The module also consists of the compression part of *backward* compatible algorithms such as *hdrmpeg,hdrjpeg, rate, gohdr* and *optimal* (see Chapter 3 for details). However, unlike the *non-backward* compatible algorithms, *backward* compatible algorithms require dual-loop encoding schemes where the base stream is first passed to the codec and encoded at specific quality levels which are then decoded and decompressed to create the residual streams. Finally, the residual streams at specific quality levels are then passed to the codec for encoding. Typically, the *backward* compatible algorithms require 8-bits/pixel/channel encoding which can be performed by any of the available codecs. There are no special requirements in this case.

A.3 Encoding module

As mentioned previously, there are a combination of factors which is used to decide the codec to be used for encoding purposes. This module consists of four codecs; the *JM H.264/AVC* (the reference H.264/AVC implementation) [AMT], the *HM H.265* (the reference HEVC implementation) [SS] and the corresponding optimised implementations x264 [Orga] and x265 [Orgb], respectively. Although, the computational performance (encoding time) of the reference implementations are extremely poor, they are essential as these non-optimised version include some additional features such as up to 14/bits/pixel/channel encoding which are not available in commercial codec implementations to date. However, if the encoding bitdepth requirements are within 10 bits/pixel/channel, the hardware optimised x264 or the x265 is used for computational efficiency purposes.

It is to be noted that the reference HEVC and x265 implementations were added to the framework since the state-of-the-art HEVC codec claims approximately 25% bitrate savings as compared to H.264/AVC. In addition to the codecs mentioned, the *ffmpeg* library is also included to encode the videos using legacy codecs such as MPEG4 and H.262 if required. Furthermore, if the codec used in H.264/AVC or x264, the intermediate *.yuv* file in encoded as a *.264* raw video stream. Similarly the raw video stream format is *.265* for HM H.265 or x265. However, the *ffmpeg* library, if used, allows the raw video stream to be wrapped in a container format such as *.mkv/.mp4/.mov*.

A.4 Decoding module

The default decoding module in the framework is the *ffmpeg* default decoder which decodes either the raw or container formatted video stream back to an intermediate *.yuv* file which is then passed to the decompression module of framework. However, there are certain issues with the ffmpeg decoder. If the bit-depth requirement of the algorithm is greater than 10 bits/pixel/channel then the reference decoder (a module of the reference codec) is used to decode the video stream as *ffmpeg* has no support for bit-depths greater than 10-bits. Also,

it has been observed that *.yuv* files encoded using the reference H.265 or x265 codecs suffer from banding artefacts if decoded using *ffmpeg*. Therefore, for the HEVC encoding, the framework uses the reference H.265 decoder module to convert *.265* video streams to *.yuv* files.

A.5 Decompression module

This module is essentially a part of the HDR video compression algorithms where the information from the intermediate *.yuv* file(s) are extracted and the compression process is reversed to reconstruct the output HDR frames.

A.6 Evaluation

This module consists of several energy-difference, structural and perceptual QA metrics described earlier in Chapter 4. Similar to the overall structure of the framework, new QA metrics can be added and existing ones removed as per the requirements of the work. The module also includes dedicated video metrics such as HDR-VQM. In this module, the raw video stream(s) at a specific quality level is decoded, decompressed and the reconstructed HDR frames are evaluated against the source/reference HDR frames to obtain the objective quality. The quality and the output bitrate of the raw video streams at different levels are subsequently stored as a data table which is then used to create the rate-distortion graphs as shown earlier in Chapters 6 and 7. As an objective evaluation can use a significant amount of disk space, the intermediate *.yuv* files, the raw video streams (.264/.265) and the reconstructed HDR frames are deleted to save disk space after the evaluation is completed.

A.7 Summary

This chapter provides an overview of the HDR video compression evaluation framework which was created in order to objectively evaluate multiple HDR video compression algorithms using multiple codecs against multiple QA metrics using a large HDR video database for source uncompressed sequences. The results obtained from the evaluations have been described earlier in Chapters 6 and 7. The programming language of choice in this case was MATLAB. However, the structure of the framework allows it to be implemented in any programming language as required.

Appendix B

HDR video sequence repository

THIS chapter provides a tone-mapped thumbnail (along with the overall dynamic range) of the 39 HDR video sequences used previously in Chapter 6 and Chapter 7 which also includes the six sequences used in Chapter 5.

(a) s01 - 21 stops (b) s02 - 23 stops (c) s03 - 28 stops

(d) s04 - 24 stops (e) s05 - 17 stops (f) s06 - 19 stops

(g) s07 - 23 stops (h) s08 - 23 stops (i) s09 - 23 stops

Figure B.1: HDR video sequences - part I

(a) s10 - 23 stops (b) s11 - 23 stops (c) s12 - 23 stops

(d) s13 - 23 stops (e) s14 - 23 stops (f) s15 - 23 stops

(g) s16 - 23 stops (h) s17 - 23 stops (i) s18 - 23 stops

(j) s19 - 23 stops (k) s20 - 23 stops (l) s21 - 23 stops

(m) s22 - 19 stops (n) s23 - 23 stops (o) s24 - 23 stops

(p) s25 - 23 stops (q) s26 - 14 stops (r) s27 - 23 stops

Figure B.2: HDR video sequences - part II

(a) s28 - 23 stops (b) s29 - 23 stops (c) s30 - 23 stops

(d) s31 - 23 stops (e) s32 - 23 stops (f) s33 - 23 stops

(g) s34 - 23 stops (h) s35 - 23 stops (i) s36 - 23 stops

(j) s37 - 23 stops (k) s38 - 23 stops (l) s39 - 17 stops

Figure B.3: HDR video sequences - part III

Bibliography

[AB08] AMEER S., BASIR O.: Objective image quality measure based on weber-weighted mean absolute error. In *2008 9th International Conference on Signal Processing* (Oct 2008), pp. 728–732.

[ABDD*14] AZIMI M., BANITALEBI-DEHKORDI A., DONG Y., POURAZAD M. T., NASIOPOULOS P.: Evaluating the performance of existing full-reference quality metrics on high dynamic range (HDR) video content. In *ICMSP 2014: International Conference on Multimedia Signal Processing* (2014), p. 811.

[ABO*15] AZIMI M., BOITARD R., OZTAS B., PLOUMIS S., TOHIDYPOUR H., POURAZAD M., NASIOPOULOS P.: Compression efficiency of HDR/LDR content. In *Quality of Multimedia Experience (QoMEX), 2015 Seventh International Workshop on* (May 2015), pp. 1–6.

[Ada80] ADAMS A.: The camera, the ansel adams photography series. *Little, Brown and Company 1981* (1980), 1983.

[Ada81] ADAMS A.: The negative, the ansel adams photograph series. *Little Brown and Company New York* (1981).

[Ada83] ADAMS A.: The print, the ansel adams photography series. *Little, Brown and Company 1981* (1983), 1983.

[AFR*07] AKYÜZ A. O., FLEMING R., RIECKE B. E., REINHARD E., BÜLTHOFF H. H.: Do hdr displays support ldr content?: A psychophysical evaluation. *ACM Trans. Graph. 26*, 3 (July 2007).

[AG"a] AG" A.: Overview of arri alexa cameras.

[AGb] AG S.-V.: Spherocam hdr.

[AG"c] AG" S.-V.: Spheron vr multisensor hdr video camera. `http://www.hdrv.org/HDRv.php`.

[AG"d] AG" W.: Civetta 360.

[AMS08a] AYDIN T. O., MANTIUK R., SEIDEL H.-P.: Extending quality metrics to full dynamic range images. In *Human Vision and Electronic Imaging XIII* (San Jose, USA, January 2008), Proceedings of SPIE, pp. 6806–10.

[AMS08b] AYDIN T. O., MANTIUK R., SEIDEL H.-P.: Extending quality metrics to full luminance range images. In *Electronic Imaging 2008* (2008), International Society for Optics and Photonics, pp. 68060B–68060B.

[AMT] ALEXIS MICHAEL TOURAPIS ATHANASIOS LEONTARIS K. S. G. S.: H.264/avc reference encoder. `http://iphome.hhi.de/suehring/tml/`.

[ari15] *Essential parameter values for the extended image dynamic range television (EIDRTV) system and programme production*. Tech. rep., Association of Radio Industries and Business, July 2015.

[Ash02] ASHIKHMIN M.: A tone mapping algorithm for high contrast images. In *Proceedings of the 13th Eurographics Workshop on Rendering* (Aire-la-Ville, Switzerland, Switzerland, 2002), EGRW '02, Eurographics Association, pp. 145–156.

[ASS02] AVCIBAŞ I., SANKUR B., SAYOOD K.: Statistical evaluation of image quality measures. *Journal of Electronic Imaging 11*, 2 (2002), 206–223.

[Ass03] ASSEMBLY I. R.: *Methodology for the subjective assessment of the quality of television pictures*. International Telecommunication Union, 2003.

[BADC11] BANTERLE F., ARTUSI A., DEBATTISTA K., CHALMERS A.: *Advanced high dynamic range imaging: theory and practice*. CRC Press, 2011.

[Bar92] BARTEN P. G.: Physical model for the contrast sensitivity of the human eye. In *SPIE/IS&T 1992 Symposium on Electronic Imaging: Science and Technology* (1992), International Society for Optics and Photonics, pp. 57–72.

[Bar99] BARTEN P. G.: *Contrast sensitivity of the human eye and its effects on image quality*, vol. 72. SPIE press, 1999.

[Bar03] BARTEN P. G.: Formula for the contrast sensitivity of the human eye. In *Electronic Imaging 2004* (2003), International Society for Optics and Photonics, pp. 231–238.

[BB71] BLACKWELL O. M., BLACKWELL H. R.: Ieri: Visual performance data for 156 normal observers of various ages. *Journal of the Illuminating Engineering Society 1*, 1 (1971), 3–13.

[BBQ*97] BOSI M., BRANDENBURG K., QUACKENBUSH S., FIELDER L., AKAGIRI K., FUCHS H., DIETZ M.: ISO/IEC MPEG-2 advanced audio coding. *Journal of the Audio engineering society 45*, 10 (1997), 789–814.

[BC15] BORER T., COTTON A.: A "display independent" high dynamic range television system.

[BDAPN14] BANITALEBI-DEHKORDI A., AZIMI M., POURAZAD M. T., NASIOPOULOS P.: Compression of high dynamic range video using the HEVC and H.264/AVC standards. In *Heterogeneous Networking for Quality, Reliability, Security and Robustness (QShine), 2014 10th International Conference on* (Aug 2014), pp. 8–12.

[Bel] BELLARD F.: The ffmpeg audio-video codec library. `https://ffmpeg.org/`.

[Bla81] BLACKWELL H.: An analytical model for describing the influence of lighting parameters upon visual performance, volume 1: Technical foundations. *Comission Internationale De LâĂŹEclairage 11* (1981).

[BLDC06] BANTERLE F., LEDDA P., DEBATTISTA K., CHALMERS A.: Inverse tone mapping. In *Proceedings of the 4th International Conference on Computer Graphics and Interactive Techniques in Australasia and Southeast Asia* (New York, NY, USA, 2006), GRAPHITE '06, ACM, pp. 349–356.

[BMP15] BOITARD R., MANTIUK R. K., POULI T.: Evaluation of color encodings for high dynamic range pixels. vol. 9394, pp. 93941K–93941K–9.

[Bod73] BODMANN H. W.: Visibility assessment in lighting engineering. *Journal of the Illuminating Engineering Society 2*, 4 (1973), 437–444.

[ČAMS11] ČADÍK M., AYDIN T. O., MYSZKOWSKI K., SEIDEL H.-P.: On evaluation of video quality metrics: an HDR dataset for computer graphics applications. In *Human Vision and Electronic Imaging XVI* (2011), Rogowitz B. E., Pappas T. N., (Eds.), vol. 7865, SPIE.

[CEB*10] CHALMERS A., EDWARDS G., BONNET G., BANTERLE F., ARTUSI A., DEBATTISTA K., LEDDA P.: HDR video data compression devices and methods. WO Patent App. PCT/EP2009/005,042.

[CH07] CHANDLER D. M., HEMAMI S. S.: VSNR: a wavelet-based visual signal-to-noise ratio for natural images. *IEEE transactions on image processing 16*, 9 (2007), 2284–2298.

[CHS*93] CHIU K., HERF M., SHIRLEY P., SWAMY S., WANG C., ZIMMERMAN K., ET AL.: Spatially nonuniform scaling functions for high contrast images. In *Graphics Interface* (1993), CANADIAN INFORMATION PROCESSING SOCIETY, pp. 245–245.

[Com] COMPANY" R. D. C. C.: Overview of the red epic camera. http://www.red.com/products/epic-dragon.

[ČWNA08] ČADÍK M., WIMMER M., NEUMANN L., ARTUSI A.: Evaluation of HDR tone mapping methods using essential perceptual attributes. *Computers & Graphics 32*, 3 (2008), 330–349.

[Dal92] DALY S. J.: Visible differences predictor: an algorithm for the assessment of image fidelity. vol. 1666, pp. 2–15.

[Dav63] DAVID H. A.: *The method of paired comparisons*, vol. 12. DTIC Document, 1963.

[DBRS*15] DEBATTISTA K., BASHFORD-ROGERS T., SELMANOVIĆ E., MUKHERJEE R., CHALMERS A.: Optimal exposure compression for high dynamic range content. *The Visual Computer 31*, 6-8 (2015), 1089–1099.

[Des] DESIGN" B.: Overview of the blackmagic cinema cameras. https://www.blackmagicdesign.com/products/cinemacameras.

[Des11] DESMET B.: *Professional Display Technology Overview*. Flanders Scientic Inc., 3 2011.

[DLCMM16a] DUFAUX F., LE CALLET P., MANTIUK R., MRAK M.: *High Dynamic Range Video: From Acquisition, to Display and Applications*. Academic Press, 2016.

[DLCMM16b] DUFAUX F., LE CALLET P., MANTIUK R., MRAK M.: *High Dynamic Range Video: From Acquisition, to Display and Applications*. Academic Press, 2016.

[DMAC03] DRAGO F., MYSZKOWSKI K., ANNEN T., CHIBA N.: Adaptive logarithmic mapping for displaying high contrast scenes. In *Computer Graphics Forum* (2003), vol. 22, Wiley Online Library, pp. 419–426.

[DMMS03] DRAGO F., MARTENS W. L., MYSZKOWSKI K., SEIDEL H.-P.: Perceptual evaluation of tone mapping operators. In *Proceedings of the SIGGRAPH 2003 conference on Sketches & applications in conjunction with*

the 30th annual conference on Computer graphics and interactive techniques - GRAPH '03 (New York, New York, USA, 2003), ACM Press, p. 1.

[DNP12] DONG Y., NASIOPOULOS P., POURAZAD M. T.: HDR video compression using high efficiency video coding (HEVC). *reproduction 6* (2012), 7.

[DS98] DAUBECHIES I., SWELDENS W.: Factoring wavelet transforms into lifting steps. *Journal of Fourier analysis and applications 4*, 3 (1998), 247–269.

[DVMS74] DE VALOIS R. L., MORGAN H., SNODDERLY D. M.: Psychophysical studies of monkey vision-iii. spatial luminance contrast sensitivity tests of macaque and human observers. *Vision research 14*, 1 (1974), 75–81.

[EF98a] EBNER F., FAIRCHILD M. D.: Development and testing of a color space (ipt) with improved hue uniformity. In *Color and Imaging Conference* (1998), vol. 1998, Society for Imaging Science and Technology, pp. 8–13.

[EF98b] EBNER F., FAIRCHILD M. D.: Finding constant hue surfaces in color space. In *Photonics West'98 Electronic Imaging* (1998), International Society for Optics and Photonics, pp. 107–117.

[Eng] ENGINEERING TOOLBOX: Illuminance - recommended light levels. `http://www.engineeringtoolbox.com/light-level-rooms-d_708.html`.

[EWMU13] EILERTSEN G., WANAT R., MANTIUK R. K., UNGER J.: Evaluation of Tone Mapping Operators for HDR-Video. *Computer Graphics Forum 32*, 7 (Oct. 2013), 275–284.

[Fai13a] FAIRCHILD M. D.: *Color appearance models*. John Wiley & Sons, 2013.

[Fai13b] FAIRCHILD M. D.: *Color appearance models*. John Wiley & Sons, 2013.

[FD81] FREEDMAN D., DIACONIS P.: On the histogram as a density estimator: L 2 theory. *Probability theory and related fields 57*, 4 (1981), 453–476.

[FFH*16] FRANÇOIS E., FOGG C., HE Y., LI X., LUTHRA A., SEGALL A.: High dynamic range and wide color gamut video coding in hevc: Status and potential future enhancements. *IEEE Transactions on Circuits and Systems for Video Technology 26*, 1 (Jan 2016), 63–75.

[FK13] FLORIAN KAINZ R. B.: *Technical Introduction to OpenEXR*. Tech. rep., Industrial Light & Magic, 2013.

[FLW02] FATTAL R., LISCHINSKI D., WERMAN M.: Gradient domain high dynamic range compression. In *Proceedings of the 29th annual conference on Computer graphics and interactive techniques* (New York, NY, USA, 2002), SIGGRAPH '02, ACM, pp. 249–256.

[FPSG96] FERWERDA J. A., PATTANAIK S. N., SHIRLEY P., GREENBERG D. P.: A model of visual adaptation for realistic image synthesis. In *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques* (1996), ACM, pp. 249–258.

[GT61] GULLIKSEN H., TUCKER L. R.: A general procedure for obtaining paired comparisons from multiple rank orders. *Psychometrika 26*, 2 (1961), 173–183.

[GT11] GARBAS J.-U., THOMA H.: Temporally coherent luminance-to-luma mapping for high dynamic range video coding with H.264/AVC. In *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on* (May 2011), pp. 829–832.

[H*52] HUFFMAN D. A., ET AL.: A method for the construction of minimum-redundancy codes. *Proceedings of the IRE 40*, 9 (1952), 1098–1101.

[HB95] HUNG P.-C., BERNS R. S.: Determination of constant hue loci for a crt gamut and their predictions using color appearance spaces. *Color Research & Application 20*, 5 (1995), 285–295.

[HBK*14] HANHART P., BERNARDO M., KORSHUNOV P., PEREIRA M., PINHEIRO A., EBRAHIMI T.: HDR image compression: A new challenge for objective quality metrics. In *Quality of Multimedia Experience (QoMEX), 2014 Sixth International Workshop on* (Sept 2014), pp. 159–164.

[HBP*15] HANHART P., BERNARDO M. V., PEREIRA M., G. PINHEIRO A. M., EBRAHIMI T.: Benchmarking of objective quality metrics for hdr image quality assessment. *EURASIP Journal on Image and Video Processing 2015*, 1 (2015), 39.

[HKE*15] HANHART P., KORSHUNOV P., EBRAHIMI T., THOMAS Y., HOFFMANN H.: Subjective quality evaluation of high dynamic range video and display for future tv. *SMPTE Motion Imaging Journal 124*, 4 (May 2015), 1–6.

[HP85] HUNT R., POINTER M.: A colour-appearance transform for the cie 1931 standard colorimetric observer. *Color Research & Application 10*, 3 (1985), 165–179.

[HR84] HELANDER M., RUPP B.: An overview of standards and guidelines for visual display terminals. *Applied Ergonomics 15*, 3 (1984), 185 – 195.

[HRE15] HANHART P., RERABEK M., EBRAHIMI T.: Towards high dynamic range extensions of HEVC: subjective evaluation of potential coding technologies. In *SPIE Optical Engineering+ Applications* (2015).

[Hun91] HUNT R.: Revised colour-appearance model for related and unrelated colours. *Color Research & Application 16*, 3 (1991), 146–165.

[Hun05a] HUNT R. W. G.: *The reproduction of colour*. Wiley. com, 2005.

[Hun05b] HUNT R. W. G.: *The reproduction of colour*. John Wiley & Sons, 2005.

[HW10] HIRAKAWA K., WOLFE P.: Optimal exposure control for high dynamic range imaging. In *Image Processing (ICIP), 2010 17th IEEE International Conference on* (Sept 2010), pp. 3137–3140.

[ic] IMS CHIPS: *HDRC Camcube*.

[IIJ94] "ITU-T, ISO/IEC, JTC 1": Generic coding of moving pictures and associated audio information - part 2: Video. *ISO/IEC* (1994).

[Int02] INTERNATIONAL TELECOMMUNICATION UNION: *Parameter values for the HDTV standards for production and international programme exchange*, recommendation itu-r bt.709-5 ed., 04 2002. BT Series.

[IR] ITU-R R. B.: 601-5: Studio encoding parameters of digital television for standard 4: 3 and wide. *Screen 16*, 9.

[ITU12] ITU: *Recommendation ITU-R BT.500-13: Methodology for the subjective assessment of the quality of television pictures*. Tech. rep., International Telecommunication Union, 2012.

[KD12] KOZ A., DUFAUX F.: A comparative survey on high dynamic range video compression. In *SPIE Optical Engineering+ Applications* (2012), International Society for Optics and Photonics, pp. 84990E–84990E.

[Kel77] KELLY D.: Visual contrast sensitivity. *Journal of modern optics 24*, 2 (1977), 107–129.

[Ken48] KENDALL M. G.: Rank correlation methods.

[KJF07] KUANG J., JOHNSON G. M., FAIRCHILD M. D.: icam06: A refined image appearance model for {HDR} image rendering. *Journal of Visual Communication and Image Representation 18*, 5 (2007), 406 – 414. Special issue on High Dynamic Range Imaging.

[KMS05] KRAWCZYK G., MYSZKOWSKI K., SEIDEL H.-P.: Lightness perception in tone reproduction for high dynamic range images. *Computer Graphics Forum 24*, 3 (2005), 635–645.

[Koe02] KOENEN R.: Overview of the mpeg-4 standard. *ISO/IEC JTC1/SC29/WG11 N 1730* (2002), 11–13.

[KP99] KIM D., PARK W.: In-plane switching liquid crystal display having high aperture ratio, may 1999. US Patent 5,907,379.

[KRTT12] KISER C., REINHARD E., TOCCI M., TOCCI N.: Real time automated tone mapping system for hdr video. In *IEEE International Conference on Image Processing* (2012), pp. 2749–2752.

[KYJF04] KUANG J., YAMAGUCHI H., JOHNSON G. M., FAIRCHILD M. D.: Testing HDR image rendering algorithms. In *Color and Imaging Conference* (2004), vol. 2004, Society for Imaging Science and Technology, pp. 315–320.

[Lar98] LARSON G. W.: Logluv encoding for full-gamut, high-dynamic range images. *Journal of Graphics Tools 3*, 1 (1998), 15–31.

[LCTS05] LEDDA P., CHALMERS A., TROSCIANKO T., SEETZEN H.: Evaluation of tone mapping operators using a high dynamic range display. *ACM Trans. Graph. 24*, 3 (jul 2005), 640–648.

[LFH15] LUTHRA A., FRANCOIS E., HUSAK W.: Call for evidence (CfE) for HDR and WCG video coding.

[LFUS06] LISCHINSKI D., FARBMAN Z., UYTTENDAELE M., SZELISKI R.: Interactive local adjustment of tonal values. *ACM Transactions on Graphics (TOG) 25*, 3 (2006), 646–653.

[LG91] LE GALL D.: Mpeg: A video compression standard for multimedia applications. *Communications of the ACM 34*, 4 (1991), 46–58.

[LK07] LEE C., KIM C.-S.: Gradient domain tone mapping of high dynamic range videos. In *Image Processing, 2007. ICIP 2007. IEEE International Conference on* (2007), vol. 3, pp. III – 461–III – 464.

[LK08] LEE C., KIM C.-S.: Rate-distortion optimized compression of high dynamic range videos. In *Proceedings of the 16th European Signal Processing Conference* (2008).

[LK12] LEE C., KIM C.-S.: Rate-distortion optimized layered coding of high dynamic range videos. *Journal of Visual Communication and Image Representation 23*, 6 (2012), 908–923.

[LRP97] LARSON G., RUSHMEIER H., PIATKO C.: A visibility matching tone reproduction operator for high dynamic range scenes. *Visualization and Computer Graphics, IEEE Transactions on 3*, 4 (Oct 1997), 291–306.

[LSC04] LEDDA P., SANTOS L. P., CHALMERS A.: A local model of eye adaptation for high dynamic range images. In *Proceedings of the 3rd International Conference on Computer Graphics, Virtual Reality, Visualisation and Interaction in Africa* (New York, NY, USA, 2004), AFRIGRAPH '04, ACM, pp. 151–160.

[Man06] MANTIUK R.: *High-Fidelity Imaging*. PhD thesis, Universität des Saarlandes, Germany, 12 2006.

[MBDC14] MELO M., BESSA M., DEBATTISTA K., CHALMERS A.: Evaluation of HDR video tone mapping for mobile devices. *Signal Processing: Image Communication 29*, 2 (Feb. 2014), 247–256.

[MDBR*16a] MUKHERJEE R., DEBATTISTA K., BASHFORD-ROGERS T., VANGORP P., MANTIUK R., BESSA M., WATERFIELD B., CHALMERS A.: Objective and subjective evaluation of high dynamic range video compression. *Signal Processing: Image Communication 47* (2016), 426 – 437.

[MDBR*16b] MUKHERJEE R., DEBATTISTA K., BASHFORD-ROGERS T., WATERFIELD B., CHALMERS A.: A study on user preference of high dynamic range over low dynamic range video. *Vis. Comput. 32*, 6-8 (June 2016), 825–834.

[MDG08] MUSTRA M., DELAC K., GRGIC M.: Overview of the dicom standard. In *ELMAR, 2008. 50th International Symposium* (Sept 2008), vol. 1, pp. 39–44.

[MDK08] MANTIUK R., DALY S., KEROFSKY L.: Display adaptive tone mapping. *ACM Trans. Graph. 27*, 3 (Aug. 2008), 68:1–68:10.

[MEMS06] MANTIUK R., EFREMOV A., MYSZKOWSKI K., SEIDEL H.-P.: Backward compatible high dynamic range mpeg video compression. *ACM Trans. Graph. 25*, 3 (July 2006), 713–723.

[MFF14] MANTEL C., FERCHIU S., FORCHHAMMER S.: Comparing subjective and objective quality assessment of HDR images compressed with JPEG-

XT. In *Multimedia Signal Processing (MMSP), 2014 IEEE 16th International Workshop on* (Sept 2014), pp. 1–6.

[MFH*02] MORONEY N., FAIRCHILD M. D., HUNT R. W., LI C., LUO M. R., NEWMAN T.: The CIECAM02 color appearance model. In *Color and Imaging Conference* (2002), vol. 2002, Society for Imaging Science and Technology, pp. 23–27.

[MGBL15] MINOO K., GU Z., BAYLON D., LUTHRA A.: On metrics for objective and subjective evaluation of high dynamic range video. vol. 9599, pp. 95990F–95990F–14.

[MKMS04] MANTIUK R., KRAWCZYK G., MYSZKOWSKI K., SEIDEL H.-P.: Perception-motivated high dynamic range video encoding. In *ACM SIGGRAPH 2004 Papers* (New York, NY, USA, 2004), SIGGRAPH '04, ACM, pp. 733–741.

[MKRH11] MANTIUK R., KIM K. J., REMPEL A. G., HEIDRICH W.: HDR-VDP-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions. *ACM Trans. Graph. 30*, 4 (July 2011), 40:1–40:14.

[MKVR09] MERTENS T., KAUTZ J., VAN REETH F.: Exposure fusion: A simple and practical alternative to high dynamic range photography. *Computer Graphics Forum 28*, 1 (2009), 161–171.

[MMS99] MANTIUK R. K., MYSZKOWSKI K., SEIDEL H.-P.: *High Dynamic Range Imaging*. John Wiley & Sons, Inc., 1999.

[MMS04] MANTIUK R., MYSZKOWSKI K., SEIDEL H.-P.: Visible difference predicator for high dynamic range images. In *Systems, Man and Cybernetics, 2004 IEEE International Conference on* (2004), vol. 3, IEEE, pp. 2763–2769.

[MMS06] MANTIUK R., MYSZKOWSKI K., SEIDEL H.-P.: Lossy compression of high dynamic range images and video. In *Electronic Imaging 2006* (2006), International Society for Optics and Photonics, pp. 60570V–60570V.

[MND13] MILLER S., NEZAMABADI M., DALY S.: Perceptual signal coding for more efficient usage of bit codes. *SMPTE Motion Imaging Journal 122*, 4 (May 2013), 52–59.

[MP94] MANN S., PICARD R.: *Beingundigital'with digital cameras*. MIT Media Lab Perceptual, 1994.

[MS06] MEYLAN L., SUSSTRUNK S.: High dynamic range image rendering with a retinex-based adaptive filter. *IEEE Transactions on image processing 15*, 9 (2006), 2820–2830.

[MT10] MOTRA A., THOMA H.: An adaptive logluv transform for high dynamic range video compression. In *2010 IEEE International Conference on Image Processing* (Sept 2010), pp. 2061–2064.

[MTM12] MANTIUK R. K., TOMASZEWSKA A., MANTIUK R.: Comparison of four subjective methods for image quality assessment. *Computer Graphics Forum 31*, 8 (2012), 2478–2491.

[NBRA83] NORMANN R. A., BAXTER B. S., RAVINDRA H., ANDERTON P. J.: Photoreceptor contributions to contrast sensitivity: applications in radiological diagnosis. *IEEE transactions on systems, man, and cybernetics*, 5 (1983), 944–953.

[NHTS87] NAYATANI Y., HASHIMOTO K., TAKAHAMA K., SOBAGAKI H.: A nonlinear color-appearance model using estévez-hunt-pointer primaries. *Color Research & Application 12*, 5 (1987), 231–242.

[NMDSLC15] NARWARIA M., MANTIUK R. K., DA SILVA M. P., LE CALLET P.: HDR-VDP-2.2: a calibrated method for objective quality prediction of high-dynamic range and standard images. *Journal of Electronic Imaging 24*, 1 (2015), 010501.

[NPDSLC15] NARWARIA M., PERREIRA DA SILVA M., LE CALLET P.: Study of high dynamic range video quality assessment. vol. 9599, pp. 95990V–95990V–13.

[NR05] NOBORU O., ROBERTSON A.: 3.9: Standard and supplementary illuminants, colorimetry, 2005.

[NSC15] NARWARIA M., SILVA M. P. D., CALLET P. L.: HDR-VQM: An objective quality measure for high dynamic range video. *Signal Processing: Image Communication 35* (2015), 46 – 60.

[Orga] ORGANIZATION" V.: The x264 video codec. `http://www.videolan.org/developers/x264.html`.

[Orgb] ORGANIZATION" V.: The x265 video codec. `https://www.videolan.org/developers/x265.html`.

[OSS*12] OHM J.-R., SULLIVAN G. J., SCHWARZ H., TAN T. K., WIEGAND T.: Comparison of the coding efficiency of video coding standardsâĂŤincluding high efficiency video coding (hevc). *IEEE Transactions on Circuits and Systems for Video Technology 22*, 12 (2012), 1669–1684.

[Pan] PANOSCAN: Panoscan mk-iii digital. `http://www.panoscan.com/MK3/`.

[PD09] PARIS S., DURAND F.: A fast approximation of the bilateral filter using a signal processing approach. *International Journal of Computer Vision 81*, 1 (2009), 24–52.

[PDAN12] POURAZAD M. T., DOUTRE C., AZIMI M., NASIOPOULOS P.: Hevc: The new gold standard for video compression: How does hevc compare with h. 264/avc? *IEEE consumer electronics magazine 1*, 3 (2012), 36–46.

[PFFG98] PATTANAIK S. N., FERWERDA J. A., FAIRCHILD M. D., GREENBERG D. P.: A multiscale model of adaptation and spatial vision for realistic image display. In *Proceedings of the 25th Annual Conference on Computer Graphics and Interactive Techniques* (New York, NY, USA, 1998), SIGGRAPH '98, ACM, pp. 287–298.

[Pho] PHOTO RESEARCH INC.: *PR-655, PR-670, PR-680 and PR-680L SpectraScanÂő Spectroradiometers*.

[PLZ*09] PONOMARENKO N., LUKIN V., ZELENSKY A., EGIAZARIAN K., CARLI M., BATTISTI F.: Tid2008-a database for evaluation of full-reference visual quality assessment metrics. *Advances of Modern Radioelectronics 10*, 4 (2009), 30–45.

[Pou87] POUNTAIN D.: Run-length encoding. *Byte 12*, 6 (1987), 317–319.

[PQ14] SMPTE standard - high dynamic range electro-optical transfer function of mastering reference displays. *SMPTE ST 2084:2014* (Aug 2014), 1–14.

[Rec12] RECOMMENDATION I.-R.: BT-2020 âĂIJparameter values for ultra-high definition television systems for production and international programme exchangeâĂİ. *International Telecommunication Union, Geneva* (2012).

[RHD*10] REINHARD E., HEIDRICH W., DEBEVEC P., PATTANAIK S., WARD G., MYSZKOWSKI K.: *High dynamic range imaging: acquisition, display, and image-based lighting*. Morgan Kaufmann, 2010.

[RHKE15] RERABEK M., HANHART P., KORSHUNOV P., EBRAHIMI T.: Subjective and objective evaluation of hdr video compression. In *9th International*

Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM) (2015), no. EPFL-CONF-203874.

[Ric11] RICHARDSON I. E.: *The H.264 advanced video compression standard*. John Wiley & Sons, 2011.

[RKAJ08] REINHARD E., KHAN E. A., AKYZ A. O., JOHNSON G. M.: *Color imaging: fundamentals and applications*. AK Peters, Ltd., 2008.

[RSSF02] REINHARD E., STARK M., SHIRLEY P., FERWERDA J.: Photographic tone reproduction for digital images. *ACM Trans. Graph. 21*, 3 (jul 2002), 267–276.

[RTS*07] REMPEL A. G., TRENTACOSTE M., SEETZEN H., YOUNG H. D., HEIDRICH W., WHITEHEAD L., WARD G.: Ldr2hdr: On-the-fly reverse tone mapping of legacy video and photographs. *ACM Trans. Graph. 26*, 3 (July 2007).

[SB59] STILES W., BURCH J.: N.p.l. colour-matching investigation: Final report (1958). *Optica Acta: International Journal of Optics 6*, 1 (1959), 1–26.

[SB06] SHEIKH H., BOVIK A.: Image information and visual quality. *Image Processing, IEEE Transactions on 15*, 2 (Feb 2006), 430–444.

[SB09] SESHADRINATHAN K., BOVIK A. C.: Motion-based perceptual quality assessment of video. In *IS&T/SPIE Electronic Imaging* (2009), International Society for Optics and Photonics, pp. 72400X–72400X.

[Sch95a] SCHLICK C.: Quantization techniques for visualization of high dynamic range pictures. In *Photorealistic Rendering Techniques*, Sakas G., MÃijller S., Shirley P., (Eds.), Focus on Computer Graphics. Springer Berlin Heidelberg, 1995, pp. 7–20.

[Sch95b] SCHLICK C.: Quantization techniques for visualization of high dynamic range pictures. In *Photorealistic Rendering Techniques*. Springer, 1995, pp. 7–20.

[Sch96] SCHEUNDERS P.: A genetic lloyd-max image quantization algorithm. *Pattern Recognition Letters 17*, 5 (1996), 547–556.

[Sel13] SELMANOVIĆ E.: *Stereoscopic high dynamic range imaging*. PhD thesis, University of Warwick, 2013.

[Ser11] SERIES" B.: *Reference electro-optical transfer function for flat panel displays used in HDTV studio production*. Radiocommunication sector of ITU, 03 2011.

[SF01] SILVERSTEIN D. A., FARRELL J. E.: Efficient method for paired comparison. *Journal of Electronic Imaging 10*, 2 (2001), 394–398.

[SHS*04] SEETZEN H., HEIDRICH W., STUERZLINGER W., WARD G., WHITEHEAD L., TRENTACOSTE M., GHOSH A., VOROZCOVS A.: High dynamic range display systems. In *ACM SIGGRAPH 2004 Papers* (New York, NY, USA, 2004), SIGGRAPH '04, ACM, pp. 760–768.

[SIMa] SIM2 MULTIMEDIA: SIM2 HDR47. `http://www.sim2.com/HDR/hdrdisplay/hdr47e_s_4k`.

[SIMb] SIM2 MULTIMEDIA: SIM2 HDR47ES6MB. `http://hdr.sim2.it/hdrproducts/hdr47es6mb`.

[SK09] SPRINGER D., KAUP A.: Lossy compression of floating point high-dynamic range images using jpeg2000. vol. 7257, pp. 72570X–72570X–11.

[SOHW12] SULLIVAN G. J., OHM J.-R., HAN W.-J., WIEGAND T.: Overview of the high efficiency video coding (hevc) standard. *IEEE Transactions on circuits and systems for video technology 22*, 12 (2012), 1649–1668.

[SS] SÜHRING K., SHARMAN K.: HEVC reference encoder. `https://hevc.hhi.fraunhofer.de/`.

[SSB06a] SHEIKH H., SABIR M., BOVIK A.: A statistical evaluation of recent full reference image quality assessment algorithms. *Image Processing, IEEE Transactions on 15*, 11 (Nov 2006), 3440–3451.

[SSB06b] SHEIKH H. R., SABIR M. F., BOVIK A. C.: A statistical evaluation of recent full reference image quality assessment algorithms. *IEEE Transactions on image processing 15*, 11 (2006), 3440–3451.

[SSBC10] SESHADRINATHAN K., SOUNDARARAJAN R., BOVIK A., CORMACK L.: Study of subjective and objective quality assessment of video. *Image Processing, IEEE Transactions on 19*, 6 (June 2010), 1427–1441.

[Ste57] STEVENS S. S.: On the psychophysical law. *Psychological review 64*, 3 (1957), 153.

[Stu14] STUMP D.: *Digital cinematography: fundamentals, tools, techniques, and workflows*. CRC Press, 2014.

[Swe98] SWELDENS W.: The lifting scheme: A construction of second generation wavelets. *SIAM Journal on Mathematical Analysis 29*, 2 (1998), 511–546.

[SYD87] SEZAN M. I., YIP K.-L., DALY S. J.: Uniform perceptual quantization: Applications to digital radiography. *IEEE Transactions on Systems, Man, and Cybernetics 17*, 4 (1987), 622–634.

[Sze10] SZELISKI R.: *Computer vision: algorithms and applications*. Springer Science & Business Media, 2010.

[Tec] TECHNOLOGIES B.: Brightside dr-37-p. `http://www.bit-tech.net/hardware/2005/10/04/brightside_hdr_edr`

[TM98] TOMASI C., MANDUCHI R.: Bilateral filtering for gray and color images. In *Computer Vision, 1998. Sixth International Conference on* (Jan 1998), pp. 839–846.

[TR93] TUMBLIN J., RUSHMEIER H.: Tone reproduction for realistic images. *IEEE Computer Graphics and Applications 13*, 6 (Nov 1993), 42–48.

[TT99] TUMBLIN J., TURK G.: LCIS: A boundary hierarchy for detail-preserving contrast reduction. In *Proceedings of the 26th annual conference on Computer graphics and interactive techniques* (1999), ACM Press/Addison-Wesley Publishing Co., pp. 83–90.

[UMM*10] URBANO C., MAGALHÃČES L., MOURA J., BESSA M., MARCOS A., CHALMERS A.: Tone mapping operators on small screen devices: An evaluation study. *Computer Graphics Forum 29*, 8 (2010), 2469–2478.

[VDSLD14] VALENZISE G., DE SIMONE F., LAUGA P., DUFAUX F.: Performance evaluation of objective quality metrics for HDR image compression. In *Proc. SPIE* (2014), vol. 9217, pp. 92170C–92170C–10.

[VMV72] VAN MEETEREN A., VOS J.: Resolution and contrast sensitivity at low luminances. *Vision research 12*, 5 (1972), 825IN2–833.

[War91] WARD G.: Real pixels. *Graphics Gems II* (1991), 80–83.

[War94a] WARD G.: A contrast-based scalefactor for luminance display. *Graphics gems IV* (1994), 415–421.

[War94b] WARD G. J.: The radiance lighting simulation and rendering system. In *Proceedings of the 21st Annual Conference on Computer Graphics and Interactive Techniques* (New York, NY, USA, 1994), SIGGRAPH '94, ACM, pp. 459–472.

[War08] WARD G.: Defining dynamic range. In *ACM SIGGRAPH 2008 Classes* (New York, NY, USA, 2008), SIGGRAPH '08, ACM, pp. 30:1–30:3.

[WB02] WANG Z., BOVIK A. C.: A universal image quality index. *IEEE Signal Processing Letters 9*, 3 (March 2002), 81–84.

[WB06] WANG Z., BOVIK A. C.: Modern image quality assessment. *Synthesis Lectures on Image, Video, and Multimedia Processing 2*, 1 (2006), 1–156.

[WBSS04] WANG Z., BOVIK A. C., SHEIKH H. R., SIMONCELLI E. P.: Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing 13*, 4 (April 2004), 600–612.

[WJN*12] WARD G., JIA W., NINAN A., TEN A., WANG G.: Encoding, decoding, and representing high dynamic range images, Aug. 21 2012. US Patent 8,248,486.

[WL05] WU B.-F., LIN C.-F.: A high-performance and memory-efficient pipeline architecture for the 5/3 and 9/7 discrete wavelet transform of jpeg2000 codec. *Circuits and Systems for Video Technology, IEEE Transactions on 15*, 12 (2005), 1615–1628.

[WRM96] WEBER E. H., ROSS H. E., MURRAY D. J.: *EH Weber on the tactile senses*. Psychology Press, 1996.

[WS82] WYSZECKI G., STILES W. S.: *Color science*. Wiley New York, 1982.

[WS04] WARD G., SIMMONS M.: Subband encoding of high dynamic range imagery. In *Proceedings of the 1st Symposium on Applied Perception in Graphics and Visualization* (New York, NY, USA, 2004), APGV '04, ACM, pp. 83–90.

[WS06] WARD G., SIMMONS M.: Jpeg-hdr: A backwards-compatible, high dynamic range extension to jpeg. In *ACM SIGGRAPH 2006 Courses* (New York, NY, USA, 2006), SIGGRAPH '06, ACM.

[WSB03] WANG Z., SIMONCELLI E. P., BOVIK A. C.: Multiscale structural similarity for image quality assessment. In *Signals, Systems and Computers, 2004. Conference Record of the Thirty-Seventh Asilomar Conference on* (2003), vol. 2, Ieee, pp. 1398–1402.

[WSBL03] WIEGAND T., SULLIVAN G. J., BJONTEGAARD G., LUTHRA A.: Overview of the h. 264/avc video coding standard. *IEEE Transactions on circuits and systems for video technology 13*, 7 (2003), 560–576.

[XPH05] XU R., PATTANAIK S. N., HUGHES C. E.: High-dynamic-range still-image encoding in jpeg 2000. *IEEE Computer Graphics and Applications 25*, 6 (Nov 2005), 57–64.

[XWHL94] XU Y., WEAVER J., HEALY D.M. J., LU J.: Wavelet transform domain filters: a spatially selective noise filtration technique. *Image Processing, IEEE Transactions on 3*, 6 (1994), 747–758.

[YBMS05a] YOSHIDA A., BLANZ V., MYSZKOWSKI K., SEIDEL H.-P.: Perceptual evaluation of tone mapping operators with real-world scenes. In *Electronic Imaging 2005* (2005), International Society for Optics and Photonics, pp. 192–203.

[YBMS05b] YOSHIDA A., BLANZ V., MYSZKOWSKI K., SEIDEL H.-P.: Perceptual evaluation of tone mapping operators with real-world scenes. In *Electronic Imaging 2005* (2005), International Society for Optics and Photonics, pp. 192–203.

[YP03] YEE Y. H., PATTANAIK S.: Segmentation and adaptive assimilation for detail-preserving display of high-dynamic range images. *The Visual Computer 19*, 7-8 (2003), 457–466.

[ZRB11] ZHANG Y., REINHARD E., BULL D.: Perception-based high dynamic range video compression with optimal bit-depth transformation. In *2011 18th IEEE International Conference on Image Processing* (2011), IEEE, pp. 1321–1324.

19th August 2014

Warwick
Medical School

PRIVATE
Ratnajit Mukherjee
WMG
University of Warwick
CV4 7AL

Dear Ratnajit,

Study Title and BSREC Reference: *PSi*, REGO-2014-1016

Thank you for submitting the above-named project to the University of Warwick Biomedical and Scientific Research Ethics Committee for research ethical review.

I am pleased to advise that research ethical approval is granted.

May I take this opportunity to wish you success with the study, and to remind you that any substantial amendments require approval from the Committee before they can be implemented. Please keep a copy of the original signed version of this letter with your study documentation.

Yours sincerely

Dr David Davies
Chair
Biomedical and Scientific
Research Ethics Sub-Committee

Biomedical and Scientific
Research Ethics Sub-Committee
A010 Medical School Building
Warwick Medical School,
Coventry, CV4 7AL.
Tel: 02476-151875
Email: BSREC@Warwick.ac.uk

Medical School Building
The University of Warwick
Coventry CV4 7AL United Kingdom
Tel: +44 (0)24 7657 4880
Fax: +44 (0)24 7652 8375

THE UNIVERSITY OF
WARWICK